Thiago Trafane Oliveira Santos

Essays on firm heterogeneity and market-power-driven misallocation

Ensaios sobre heterogeneidade de firmas e má alocação de recursos devido a poder de mercado

Brasília

2024

Thiago Trafane Oliveira Santos

Essays on firm heterogeneity and market-power-driven misallocation

Ensaios sobre heterogeneidade de firmas e má alocação de recursos devido a poder de mercado

Doctoral Thesis presented to the Post-graduate Program in Economics of the Department of Economics at the University of Brasília, as a partial requirement for obtaining the degree of Doctor in Economics.

Universidade de Brasília - UnB

Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas - FACE

Departamento de Economia - ECO

Programa de Pós-Graduação

Supervisor Dr. Daniel Oliveira Cajueiro

Brasília

2024

To my daughter, Isabela, whose presence in our lives fills us with hope, love, and anticipation. Though we have yet to meet, you have already inspired me to be my best self. This thesis is dedicated to you, as a symbol of the endless possibilities that lie ahead.

Acknowledgements

First and foremost, I would like to acknowledge my family. To my wife, Wanessa, for her unwavering encouragement, support, and companionship — this work would not have been possible without her by my side. I also extend my deepest gratitude to my parents, Fabio and Rita, my sister Melissa, and my nieces Ana Clara and Priscila for their steadfast support and affection throughout this journey.

I am profoundly thankful to my academic advisor, Dr. Daniel O. Cajueiro, and my former academic advisor, Dr. Tomás R. Martinez, for their patient guidance, insightful advice, and constant availability. Their invaluable knowledge and mentorship have been instrumental in my academic journey. It has been an immense honor to be their pupil.

I sincerely thank Dr. Marcelo C. Medeiros for graciously hosting me at the University of Illinois Urbana-Champaign during the final months of my PhD. This stay was a transformative experience that had a profound impact on both my personal development and my work.

I am also grateful to all my professors at UnB for sharing their extensive knowledge with me.

Lastly, I would like to thank my friends and colleagues who have been so generous with their time and attention throughout this journey.

Abstract

This work comprises three papers that explore the interplay of firm heterogeneity and market-power-driven misallocation. The first paper develops a static Cournot model that estimates market-power-driven misallocation using essentially standard macroeconomic data. We utilize this model to revisit key facts of economic growth. On the one hand, we evaluate the world income frontier as proxied by the US, finding that changes in misallocation can significantly impact short-run growth. On the other hand, we examine the economic performance around the world. We conclude that misallocation enhances our understanding of cross-country income differences, even though a substantial unexplained portion persists. We also find a lack of convergence in allocative efficiency, suggesting market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions. The second paper also uses the Cournot model developed in the first article, but it focuses on understanding the role of misallocation in recent economic cycles in Brazil. We find an upward trend in the allocative efficiency and that the cycles in total factor productivity (TFP) are mainly due to misallocation, with the economic boom in the mid-2000s being primarily attributed to efficiency gains. The technology component of TFP grows much more steadily, around 0.8-0.9% per year, suggesting it reflects the structural characteristics of the economy. Finally, the third paper evaluates Zipf's law for the distribution of firm size by the number of employees in Brazil. We find that this "law" provides a very good, although not perfect, approximation to data. However, a lognormal distribution also performs well and even outperforms Zipf's law in certain cases.

Keywords: TFP, misallocation, Cournot model, firm size distribution, Zipf's law.

Resumo

Este trabalho é composto por três artigos que exploram a interação entre heterogeneidade de firmas e má alocação de recursos devido a poder de mercado. O primeiro artigo desenvolve um modelo estático de Cournot que estima a má alocação devido a poder de mercado utilizando essencialmente dados macroeconômicos. Nós utilizamos esse modelo para reexaminar importantes fatos do crescimento econômico. Por um lado, avaliamos o desempenho dos países mais ricos, medindo-o nos Estados Unidos. Constatamos que mudanças na eficiência alocativa podem impactar significativamente o crescimento de curto prazo. Por outro lado, examinamos o desempenho econômico ao redor do mundo. Concluímos que a má alocação desempenha um papel significativo na explicação das diferenças de renda entre países, embora uma parte substancial permaneça inexplicada. Ademais, não encontramos suporte à hipótese de convergência para a eficiência alocativa, sugerindo que a má alocação devido a poder de mercado está relacionada, no longo prazo, a fatores duradouros específicos de cada país como as instituições. O segundo artigo também utiliza o modelo de Cournot desenvolvido no primeiro, mas visando entender o papel da má alocação nos ciclos econômicos recentes do Brasil. Nós encontramos uma tendência de melhora na eficiência alocativa e que os ciclos na produtividade total dos fatores (PTF) brasileira são principalmente devidos à eficiência alocativa, com o boom econômico de meados dos anos 2000 sendo principalmente atribuído aos ganhos de eficiência. O componente tecnológico da PTF cresce de maneira muito mais estável, em torno de 0,8-0,9% ao ano, sugerindo que reflete as características estruturais da economia. Por fim, o terceiro artigo avalia a lei de Zipf para a distribuição do tamanho de empresas por pessoal ocupado no Brasil. Nós constatamos que essa "lei" fornece uma aproximação muito boa, embora não perfeita, para os dados. No entanto, a distribuição lognormal também apresenta bom desempenho e até supera a lei de Zipf em certos casos.

Palavras-chaves: PTF, má alocação, modelo de Cournot, tamanho das empresas, lei de Zipf.

List of Figures

List of Tables

List of abbreviations and acronyms

2SLS	Two-Stage Least Squares
BLUE	Best Linear Unbiased Estimator
BNDES	Banco Nacional de Desenvolvimento Econômico e Social (Brazilian Development Bank)
CEMPRE	Cadastro Central de Empresas (Central Register of Enterprises)
CES	Constant Elasticity of Substitution
CNAE	Classificação Nacional de Atividades Econômicas (National Classification of Economic Activities)
DI	Depósito Interfinanceiro (Interbank Deposit)
ERP	Equity Risk Premium
EU	European Union
FGV	Fundação Getulio Vargas (Getulio Vargas Foundation)
FOC	First-Order Condition
GDP	Gross Domestic Product
GMM	Generalized Methods of Moments
HHI	Herfindahl–Hirschman index
IBGE	Instituto Brasileiro de Geografia e Estatística (Brazilian Institute of Geography and Statistics)
IRPJ/CSLL	Imposto de Renda da Pessoa Jurídica/Contribuição Social sobre o Lucro Líquido (Corporate Income Tax)

ISIC	International Standard Industrial Classification of All Economic Activities
IT	Information Technology
KS	Kolmogorov–Smirnov
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MVUE	Minimum Variance Unbiased Estimator
NIPA	National Income and Productivity Accounts
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PNAD	Pesquisa Nacional por Amostra de Domicílios (National Household Sample Survey)
PPP	Purchasing Power Parity
PTF	Produtividade Total dos Fatores
SOC	Second-Order Condition
SSR	Sum of Squared Residuals
TFP	Total Factor Productivity
UK	United Kingdom
US	United States
WACC	Weighted Average Cost of Capital

Contents

1 Introduction

This work comprises three papers that explore the interplay of firm heterogeneity and market-power-driven misallocation. The first paper develops a static Cournot model that estimates market-power-driven misallocation using essentially standard macroeconomic data. This model enables us to assess the role played by misallocation in shaping total factor productivity (TFP), an analysis that has been hindered by constraints in the availability of firm-level data. We apply this framework to decompose aggregate TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries from the Penn World Table 10.01. Utilizing this decomposition, we revisit key facts of economic growth. On the one hand, we evaluate the world income frontier as proxied by the US, finding that changes in misallocation can significantly impact short-run growth. On the other hand, we examine the economic performance around the world. We conclude that misallocation enhances our understanding of cross-country income differences, even though a substantial unexplained portion persists. We also find a lack of convergence in allocative efficiency, suggesting market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

The second paper also uses the Cournot model developed in the first article, but it focuses on understanding the role of misallocation in recent economic cycles in Brazil. We find an upward trend in Brazilian allocative efficiency between 2000 and 2019, reflecting the observed increase in the labor income share and, thus, the estimated decrease in the average markup, in sharp contrast with most developed countries. Additionally, we find that the cycles in Brazilian TFP are mainly due to allocative efficiency, with the economic boom in the mid-2000s being primarily attributed to efficiency gains. The technology component of TFP grows much more steadily, around 0.8-0.9% per year, suggesting it reflects the structural characteristics of the economy.

Finally, the third paper addresses Zipf's law, which states that the probability of a variable being larger than s is roughly inversely proportional to s. More precisely, this paper evaluates if this “law” holds for the distribution of firm size by the number of employees in Brazil. We use publicly available binned annual data from the Central Register of Enterprises (CEMPRE), which is held by the Brazilian Institute of Geography and Statistics (IBGE) and covers all formal organizations. Remarkably, we find that Zipf’s law provides a very good, although not perfect, approximation to data for each year between 1996 and 2020 at the economy-wide level and also for agriculture, industry, and services alone. However, a lognormal distribution also performs well and even outperforms Zipf’s law in certain cases.

2 Revisiting the facts of economic growth: insights from assessing misallocation over 70 years for up to 100 countries

2.1 Introduction

Aggregate total factor productivity (TFP) stands out as a crucial determinant of economic performance across countries and time (Klenow; Rodríguez-Clare, 1997; Caselli, 2005; Caselli, 2016; Jones, 2016; Bergeaud; Cette; Lecat, 2018; Crafts; Woltjer, 2021). However, TFP is usually a residual, "a measure of our ignorance," capturing the unexplained portion of such economic outcomes. This fact has spurred extensive research into the determinants of TFP, as illustrated in the title of a paper by Prescott (1998): "Needed: A Theory of TFP." Jones (2016, p.46) argues that "[...] the literature on misallocation has emerged to provide the kind of theory that Prescott was seeking." Nevertheless, measuring misallocation typically requires firm-level data available only for selected countries in specific years – and usually restricted to manufacturing. This limitation has presented obstacles to including misallocation estimates into standard growth and development accounting exercises. In this paper, we undertake this task by decomposing aggregate TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries.

We employ a static model in which firms engage in Cournot competition. In our model, (i) a firm's market share is strictly increasing in its productivity, and (ii) a firm with greater market share faces lower price elasticity of demand, charging a higher markup. Consequently, variations in productivity among firms result in differences in markups, leading to inefficient allocation of resources as the marginal products would not be equalized across firms. In this sense, our model exclusively deals with market-power-driven misallocation. This aligns our study more closely with Edmond, Midrigan and Xu (2022), which estimates the welfare costs of markups, rather than with Hsieh and Klenow (2009), which measures generic allocative distortions through wedges. It is worth noting that this channel of allocative inefficiency is typical of Cournot models (Atkeson; Burstein, 2008). The novel aspect of our model lies in its primary reliance on standard macroeconomic data for calibration, which is precisely what enables our comprehensive assessment of misallocation across both time and space.

The minimal microdata requirement does not come without a cost. Our model is less flexible than other related oligopoly models, such as Edmond, Midrigan and Xu (2015) and Loecker, Eeckhout and Mongey (2021), in three important aspects. First, instead of having firms producing distinct goods in each sector over a continuum of sectors, we suppose there is only one sector where firms produce the same good. Second, there are no fixed costs. However, as the goods are homogeneous, a firm may still be inactive, and consequently, we also incorporate an entry stage to determine the set of active firms in equilibrium. Third, we assume free entry among low-productivity firms, supposing the number of inefficient firms is sufficiently large to the extent that some will not be active. As a result, the profit of the marginal active firm should be low, insignificant.

We show that allocative efficiency and other key model expressions do not depend on parameters such as the price elasticity of demand and the number of potential or active firms. They require only the empirical distribution of active firms' productivity, which is crucial for the empirical objective of decomposing TFP using macroeconomic data. This feature stems from the model's stronger assumptions, notably free entry with no fixed costs – a situation feasible only if goods are homogeneous. In such circumstances, the market share of the marginal active firm should be negligible, regardless of parameters values. In contrast, when goods are heterogeneous, firm selection occurs only with fixed costs, leading the marginal firm to produce positive quantities in equilibrium, even with free entry. In such cases, the output of the marginal firm is contingent upon model parameters (e.g., elasticity of demand and fixed costs), as identified by the null-profit condition.

Therefore, to decompose the TFP, we only need to estimate the empirical distribution of active firms' productivity. Consistent with a large part of the literature (Melitz; Redding, 2015; Edmond; Midrigan; Xu, 2015; Edmond; Midrigan;

Xu, 2022), we assume this distribution is Pareto, truncated within the productivity range. We test different values for the shape parameter and search the distributional support by matching (i) the aggregate TFP and (ii) the cost-weighted average of firm-level markups. To compute these moments using real-world data, we parameterized the firms' constant-returns-to-scale Cobb-Douglas production function, which has both physical and human capital as inputs. We employ the standard physical capital share parameter $\alpha = 1/3$ in our baseline calibration. Once α is determined, the two target moments are easily obtained from standard macroeconomic data. On the one hand, the TFP is backed out as a residual in the aggregate production function. On the other hand, the cost-weighted average markup equals $1 - \alpha$ divided by the labor share of national income, aligning with Hall (1988), which has recently gained widespread adoption for assessing firm-level markups (Loecker; Warzynski, 2012; Loecker; Eeckhout, 2018; Loecker; Eeckhout; Unger, 2020; Traina, 2018; Calligaris; Criscuolo; Marcolin, 2018; Autor et al., 2020).

Our estimation of allocative efficiency relies exclusively on the average markup, computed using the labor share of national income. Consequently, our oligopoly model accommodates labor income share variability, interpreting it as informative of misallocation.[1] More precisely, our model associates a higher labor share and, thus, a lower average markup with reduced misallocation. In particular, the allocation becomes optimal when the markup converges to one. This is intuitive as an average markup closer to the competitive unitary level suggests a less distorted economy, indicating closer proximity to optimal allocation. Once allocative efficiency is estimated from the average markup, we pin down the other element of productivity – the technology component – using TFP data. Hence, in our model, the residual of the production function is not the TFP itself but rather only its technology component, which is cleaner as it is free of misallocation effects.

Employing this calibration strategy, we decompose the aggregate TFP for various countries and years using data from the Penn World Table 10.01 (Feenstra; Inklaar; Timmer, 2015). We use this decomposed TFP data to revisit key facts of economic growth. On the one hand, we evaluate nations at the income frontier, using

[1] In contrast, under perfect competition, this labor share would be identical everywhere, with inputs always allocated optimally, if one continues to consider a universal Cobb-Douglas production function.

the United States as a proxy, as done in Jones (2016). We begin by demonstrating that our market-power-driven misallocation estimates are consistent with those obtained from the oligopoly model of Edmond, Midrigan and Xu (2022) for the US. This serves as an important validation of our model, particularly considering that their model is more flexible, general, requiring firm-level data for calibration. Our empirical analysis, spanning from 1954 to 2019, reveals that changes in misallocation can significantly impact short-run growth. For example, during 2000-2007, the US witnessed notable technological improvement coupled with declining allocative efficiency. This suggests that the dot-com boom and information technology (IT) advancements led to productivity gains but concentrated in certain firms. On a more general note, the technology component seems to grow more steadily than the TFP itself, around 1% per year. Notable exceptions are the periods of 1954-1973 and 2000-2007 when technology contributed approximately 2% annually.

On the other hand, we examine the economic performance around the world. Although the Penn World Table 10.01 provides data from 1950 to 2019, the sample coverage varies across the years, encompassing more than a hundred countries in recent years. Our findings underscore the significant role of misallocation in explaining cross-country income differences. Despite its significance, a considerable unexplained portion persists, still constituting the majority of observed variability in most cases. Additionally, we obtain limited support for the convergence hypothesis in income and either TFP component. Consequently, countries do not appear to be converging over time to a common degree of allocative efficiency, indicating that the level of efficiency is country-specific even in the long run. Interestingly, this suggests that market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

Related literature. From a methodological perspective, our work is related to several papers that embed oligopoly market structures in macroeconomic models. In most models, firms' decisions are static as in our own (Bernard et al., 2003; Atkeson; Burstein, 2008; Edmond; Midrigan; Xu, 2015; Loecker; Eeckhout; Mongey, 2021), but there are also models in which they are dynamic (Peters, 2020; Wang; Werning, 2022; Edmond; Midrigan; Xu, 2022).[2] Besides their methodological similarities,

[2] Berger, Herkenhoff and Mongey (2022) also include strategic behavior in a general equilibrium

these papers have very different objectives. For instance, Atkeson and Burstein (2008) seek to explain the observed deviations from relative purchasing power parity, Edmond, Midrigan and Xu (2015) evaluate the impact of opening up to trade on productivity through misallocation, and Wang and Werning (2022) study how market concentration affects the potency of monetary policy. Among them, Edmond, Midrigan and Xu (2022) is the closest to our paper in terms of purposes, as they assess the welfare costs of markups using (also) a Cournot model with free entry. However, their model is dynamic and involves a different form of free entry. While in our model firms decide to enter the market only *after* knowing their productivity, in their model firms should decide *before* knowing it. Furthermore, Edmond, Midrigan and Xu (2022) estimate such costs just for the US, not on a period-by-period basis, and considers markup costs associated with factors beyond misallocation, such as inefficient entry. Furthermore, they need firm-level data to calibrate their model. We contribute to this literature by estimating a model using minimal microdata, allowing us to assess market-power-driven misallocation for a broader set of countries and years.

Given our goal of revisiting key facts of economic growth, this study closely aligns with the development economics literature, particularly that focused on macroeconomic analyses and development accounting exercises (Klenow; Rodríguez-Clare, 1997; Prescott, 1998; Caselli, 2005; Caselli, 2016; Jones, 2016). Jones (2016) serves as our primary reference, offering a comprehensive overview of economic growth facts, some of which we reexamine in this study. Our contribution to this literature lies in providing a comprehensive assessment of misallocation across both time and space. As a consequence, we address allocative efficiency issues within development accounting frameworks for several different years. Furthermore, we provide evidence on the impacts of misallocation on frontier growth, a topic that has received considerably less attention than the study of misallocation and development (Jones, 2016).

Our work also shares common ground with the growth-accounting literature that decomposes TFP growth into technology and allocative efficiency components

framework, but they do that by considering an oligopsony model for the labor market.

(Basu; Fernald, 2002; Petrin; Levinsohn, 2012; Baqaee; Farhi, 2020).[3] It is most closely related to Baqaee and Farhi (2020) since they, like us, measure allocative efficiency as the distance from the optimal allocation.[4] However, their model is much more general as they do not impose any specific market structure and allow for arbitrary elasticities of substitution, returns to scale, factor mobility, and input-output network linkages. As a result, the quantification of their model requires extensive microdata that are hardly available, while our model requires mainly macroeconomic data.

Finally, our paper is also related to the literature on misallocation (Restuccia; Rogerson, 2008; Hsieh; Klenow, 2009; Restuccia; Rogerson, 2013). However, in this literature, misallocation is a result of exogenous wedges, whose estimation requires extensive firm-level data, similar to growth-accounting methods, especially Baqaee and Farhi (2020).[5] In contrast, misallocation is endogenous in our model, emerging as an equilibrium outcome. Furthermore, in this literature, the typical goal is to gauge the importance of misallocation in explaining cross-country TFP differences, usually within the manufacturing sector due to data availability, while we focus on economy-wide outcomes, both across countries and over time.

The remainder of the paper proceeds as follows. Section 2.2 presents the model, while the model quantification is explained in Section 2.3. Section 2.4 discusses data details and empirical results for the global frontier, proxied by the US. Section 2.5 presents similar assessments but for the economic performance

[3] Baqaee and Farhi (2020) present a review of such methods, showing that they also use different definitions for the relevant aggregate productivity. For a broader review of the growth-accounting literature, see Hulten (2010).

[4] Measuring allocative efficiency as the distance from optimal allocation aligns with the prevailing notion in the misallocation literature. However, alternative concepts exist, particularly in the growth-accounting literature. As pointed out by Baqaee and Farhi (2020, p.107), "[...] the growth-accounting notion of changes in allocative efficiency due to the reallocation of resources to more or less distorted parts of the economy over time is very different from the misallocation literature's notion of allocative efficiency measured as the distance to the Pareto-efficient frontier."

[5] Indeed, Baqaee and Farhi (2020, p.107) "[...] provide an analytical formula for the social cost of distortions, generalizing misallocation formulas like those of Hsieh and Klenow (2009) to economies with arbitrary input-output network linkages, numbers of factors, microeconomic elasticities of substitution, and distributions of distorting wedges."

around the world. Section 2.6 briefly comments on two model extensions. Finally, Section 2.7 concludes.

2.2 Model

We present two versions of the model. In the first version, as usual in oligopoly models, we assume there is a discrete number of firms. A shortcoming of this version is that some key results are only *approximately* valid as marginal adjustments in the number of firms are not allowed. In the second version, we consider a continuum of firms, when the standard practice is to assume these null-measure firms ignore the impacts of their decisions on aggregate outcomes even though they exist. We follow a different approach and suppose firms pay at least some attention to them. Under this assumption, we get essentially the same results as the discrete model but holding *exactly*.

In this section, we specifically emphasize the economic content of the model. For step-by-step derivation and formal proofs, please refer to Appendix 2.A for the discrete version and Appendix 2.B for the continuous version.

2.2.1 Discrete number of firms

We refrain from delving into households' behavior, as it is essentially irrelevant to our results; all we need is to assume that more consumption is always preferred to less. Consequently, the model focuses solely on the firms' side, where misallocation originates. Additionally, since firms' decisions are static in our model, we suppress the time subscript for notational simplicity.

Environment and technology. In a closed economy, N potential entrant firms produce a single good. The price elasticity of demand for this good is strictly negative, with its absolute value denoted by η, where $1 < \eta < \infty$. One may consider that there are several different goods, each produced within distinct sectors, but which can be represented by a single-sector (or single-good) economy. In this interpretation, explored in Appendix 2.C, our model would apply to this representative sector, with η being equal to the elasticity of substitution across sectors' goods.

Since firms' goods are homogeneous, the aggregate output Y is

$$Y \equiv \sum_{i=1}^{N} Y_i \tag{2.1}$$

being Y_i the production of firm i, which is given by the Cobb-Douglas function

$$Y_i = A_i K_i^{\alpha} H_i^{1-\alpha} \tag{2.2}$$

where $K_i \geq 0$ is the stock of physical capital, $H_i \geq 0$ is the stock of human capital, and $A_i > 0$ is a productivity parameter, all for firm i, while $\alpha \in (0,1)$. In the following, let $\underline{\underline{A}} \equiv \min_i\{A_i\}$ and $\overline{A} \equiv \max_i\{A_i\}$ be the technology frontier of this economy, with $0 < \underline{\underline{A}} < \overline{A} < +\infty$.

Market competition and optimal decision. Firms engage in Cournot competition, meaning each firm chooses its output taking as given the output chosen by the other firms in the economy, as well as the wage $w > 0$ and the rental cost of physical capital $r > 0$. There are no fixed costs. Formally, each firm $i \in \{1, 2, ..., N\}$ solves the profit maximization problem

$$\begin{aligned} \max_{Y_i} \quad & pY_i - wH_i - rK_i = (p - MC_i)\, Y_i \\ \text{s.t.} \quad & w > 0, r > 0, p = p(Y), Y_j \geq 0 \ \forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{aligned} \tag{2.3}$$

where p is the price of the good and $MC_i = \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{A_i}$ is the Cobb-Douglas marginal cost of firm i. The price is given by the inverse demand function $p(Y)$, with $-\left(\frac{\partial p}{\partial Y}\frac{Y}{p}\right)^{-1} \equiv \eta$. Solving the First-Order Condition (FOC) of this optimization problem,

$$p = MC_i \frac{\eta_i}{\eta_i - 1} \tag{2.4}$$

where $\eta_i = \frac{\eta}{s_i} > 1$ is the price elasticity of demand faced by a firm with market share $s_i \equiv Y_i/Y$. Thus, the markup of firm i is $\mu_i \equiv \frac{\eta_i}{\eta_i - 1} = \left(1 - \frac{s_i}{\eta}\right)^{-1}$. Since the Second-Order Condition (SOC) for profit maximization also holds under $1 < \eta < \infty$, Equation (2.4) represents firm i optimal decision as long as $\mu_i \geq 1 \leftrightarrow s_i \geq 0$, that is, for every active firm i.

Equilibrium allocation. Using (2.4) for any active firms i and j, one can show that

$$s_i - s_j = \left(1 - \frac{MC_i}{MC_j}\right)(\eta - s_j) = \left(1 - \frac{A_j}{A_i}\right)(\eta - s_j) \tag{2.5}$$

implying that a more productive firm will have a higher market share, since $A_j > A_i \to s_j > s_i$ as $\eta > 1 \geq s_j$. Moreover, for A_i sufficiently close to zero, $s_i < 0$, implying that firms with very low productivity could not be active.[6] Thus, even though there are no fixed costs, we need to consider an entry stage to obtain the set of active firms in equilibrium.

We seek an equilibrium in which (i) each active firm i has non-negative profits (or, equivalently, $\mu_i \geq 1$) and (ii) non-active firms would make strictly negative profits if they entered the market. However, this equilibrium is usually not unique (Atkeson; Burstein, 2008; Edmond; Midrigan; Xu, 2015; Loecker; Eeckhout; Mongey, 2021). To avoid multiple equilibria, we discard any equilibrium in which a non-active firm has a strictly lower marginal cost than an active firm. As a consequence, in the equilibrium, a firm with productivity $\underline{A}$ will serve as the cutoff for active firms, such that firm i is active if and only if $A_i \geq \underline{A}$.[7] To identify this set of active firms, one can employ either of the two algorithms utilized by Atkeson and Burstein (2008), Edmond, Midrigan and Xu (2015), and Loecker, Eeckhout and Mongey (2021): (i) add firms one by one in descending order of productivity or (ii) assume all firms are initially in the market and drop one by one the firm with the lowest negative profit. In fact, any refinement device providing such a set

[6] To see that, rewrite (2.5) as $s_i = \eta - (\eta - s_j)\frac{A_j}{A_i} \leq \eta - (\eta - 1)\frac{A_j}{A_i}$ as $s_j \leq 1$ and use the fact that $\eta > 1$.

[7] Strictly speaking, the statement should be that firm i is active if $A_i > \underline{A}$ and only if $A_i \geq \underline{A}$, as it is possible to encounter situations where not all firms with productivity $\underline{A}$ are active. However, for the sake of clarity and convenience, we opt to rule out such possibilities, which results in just a low loss of generality. After all, in cases where such situations arise, we can lower the productivity of the inactive firms with productivity $\underline{A}$ by an arbitrarily small value. This adjustment leads to (essentially) the same model but with the advantage of satisfying the stronger statement of the main text.

of active firms should yield the same result, meaning that this refined equilibrium is unique.[8]

The entry stage ensures free exit. We also ensure free entry among low-productivity firms, supposing inefficient technologies are common knowledge as we assume there are firms in any neighborhood of the null productivity (so, $\underline{\underline{A}} \to 0$ and $N \to \infty$). In such circumstances, (an infinite number of) sufficiently low-productivity firms will not be active from what we have seen in (2.5). This implies the profit of the marginal active firm, the firm with productivity $\underline{A}$, should be relatively low since, otherwise, the entry of new firms would become profitable. In this context, we consider that the profit of this marginal firm is approximately null. Formally, for $\underline{A} = A_j$, $[p - MC_j]\, Y_j = MC_j\,(\mu_j - 1)\, Y_j \approx 0$, which holds if and only if $s_j \approx 0$ as $\mu_j = (1 - s_j/\eta)^{-1}$. Hence, our assumption of free entry implies that the market share of the marginal active firm should remain close to zero, regardless of parameters values. This characteristic will be crucial for our empirical strategy and stems from the model's stronger assumptions. Notably, the free entry condition in an environment where firms are not subject to fixed costs – a situation made feasible only due to the homogeneity of firms' goods. In contrast, when goods are heterogeneous, firm selection occurs only with fixed costs, leading the marginal firm to produce positive quantities in equilibrium even with free entry. In such cases, the output of the marginal firm is contingent upon model parameters (e.g., elasticity of demand and fixed costs), as identified by the near-zero-profit condition.

Evaluating Equation (2.5) for $A_j = \underline{A}$ and thus $s_j \approx 0$, it is easy to see that

$$s(A_i) \equiv s_i \approx \eta\,(1 - \underline{A}/A_i) \tag{2.6}$$

for $A_i \geq \underline{A}$, which clearly shows that s_i strictly increases with A_i. Moreover, summing Equation (2.5) in j over all active firms and using $s(\underline{A}) \approx 0$ one more time, one can show that

$$\eta \approx \frac{1}{N_a\,[1 - \mathrm{E}_a\,(\underline{A}/A)]} \tag{2.7}$$

[8] To show that, assume by contradiction two refinement devices provide different such sets of active firms. Thus, necessarily, one device (say, the first) should not include some of the lowest-productivity active firms of the other (say, the second). Since the set of active firms from the first device represents an equilibrium, including these missing firms would yield negative profits for at least one active firm, contradicting the results from the second device.

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|A \geq \underline{A}\right) = \sum_{A \geq \underline{A}} h(A) \frac{g(A)}{1-G(\underline{A})}$ is the expected value of a function h over active firms under the empirical distribution, $g(A)$ is the empirical probability of A, $G(A) = \sum_{a<A} g(a)$ is the empirical cumulative distribution function, and $N_a \equiv N\left(1 - G(\underline{A})\right)$ is the number of active firms.

Aggregate productivity and misallocation. From Equations (2.1) and (2.2), note

$$Y = \sum_{i=1}^{N} A_i K_i^{\alpha} H_i^{1-\alpha} = \overline{A}\Omega K^{\alpha} H^{1-\alpha} \tag{2.8}$$

where $K \equiv \sum_{i=1}^{N} K_i$, $H \equiv \sum_{i=1}^{N} H_i$, and $\Omega \equiv \sum_{i=1}^{N} \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right)$, with $\theta_{Ki} \equiv \frac{K_i}{K}$ and $\theta_{Hi} \equiv \frac{H_i}{H}$. As firms produce homogeneous goods, the optimal allocation entails assigning all inputs to the most productive firm, when $\Omega = 1$ and $Y = \overline{A}K^{\alpha}H^{1-\alpha}$. Hence, $\Omega \in (0, 1]$ gauges the distance of aggregate TFP $\overline{A}\Omega$ from its optimal level, $\overline{A}$, being our measure of allocative efficiency.

Since each firm uses its inputs optimally, taking the same inputs' rental prices as given, every active firm i chooses the same physical-to-human capital ratio, and thus $\theta_{Ki} = \theta_{Hi} \equiv \theta_i$. As a result,

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i K_i^{\alpha} H_i^{1-\alpha}}{\overline{A}\Omega K^{\alpha} H^{1-\alpha}} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{2.9}$$

$$\overline{A}\Omega = \frac{1}{\sum_{i=1}^{N} \frac{s(A_i)}{A_i}} \tag{2.10}$$

where we use (2.2) and (2.8) to get (2.9), while (2.10) is obtained by summing (2.9) over all firms. Therefore, similarly to Edmond, Midrigan and Xu (2015) and Edmond, Midrigan and Xu (2022), aggregate productivity is a quantity-weighted harmonic mean of firms' productivity.

Finally, by plugging Equations (2.6) and (2.7) into (2.10), we get

$$\Omega \approx \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{A}/A)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{A}/A)\right]} \tag{2.11}$$

which has some interesting properties. First, Ω improves when $\underline{A}$ increases due to the exit of less productive active firms from the market (Appendix 2.A.4), suggesting that allocative efficiency is closely related to productivity dispersion among active firms. Second, since with no productivity dispersion, any allocation of resources is optimal, $\Omega \to 1$ when $\underline{A} \to \overline{A}$.[9] Note that, in this limit case, the market would be in perfect competition since all active firms have unitary markups and null profits, with $N_a \to \infty$. Interestingly, even though there are infinite potential entrant firms, the number of active firms would be infinite only in this particular circumstance. After all, from Equation (2.7), $N_a \to \infty$ if and only if $\mathrm{E}_a\left(\underline{A}/A\right) \to 1$, which can occur only if there is an infinite number of firms with productivity around $\overline{A}$.

Average markup. Using this model, we can also compute the cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N}\left(\frac{H_i w + K_i r}{Hw+Kr}\right)\mu_i = \sum_{i=1}^{N}\theta_i\mu_i$ through

$$\mu \approx \frac{\overline{A}\Omega}{\underline{A}} \tag{2.12}$$

As we discuss in Section 2.3, this expression is crucial for our empirical strategy.

2.2.2 Continuum of firms

With a discrete number of firms, some equations such as (2.11) are only *approximately* valid. This happens because $s(\underline{A}) > 0$ could occur in equilibrium even with free entry since the fact that all active firms are making strictly positive profits does not necessarily mean that it would be profitable for a non-active (discrete) firm to enter the market. Thus, in the discrete case, one can only argue that the profit of the least productive active firm is, in some sense, low — something we incorporate when using $s(\underline{A}) \approx 0$. To get equations that are *exactly* valid, we need to work with a continuum of firms. However, under such circumstances, the standard practice is to assume these null-measure firms ignore the impacts of their decisions on aggregate outcomes even though they exist. For instance, in models of monopolistic competition, each intermediate firm ignores its impact on aggregate output, but it exists as the optimal decision of final good producers relies on the marginal effects on aggregate output of employing more intermediate goods.[10]

[9] This result relies solely on the assumption of homogeneous goods, not depending on the Cournot model. Please refer to Appendix 2.A.4 for formal proof.

[10] For the standard CES aggregator $Y = \left(\int_0^1 Y_j^{\frac{\sigma-1}{\sigma}}dj\right)^{\frac{\sigma}{\sigma-1}}$, $\partial Y/\partial Y_i = \left(\int_0^1 Y_j^{\frac{\sigma-1}{\sigma}}dj\right)^{\frac{1}{\sigma-1}} Y_i^{-\frac{1}{\sigma}} > 0$, with $\partial Y/\partial Y_i = 1$ for homogeneous goods ($\sigma \to +\infty$).

Similarly, in oligopoly models based on Atkeson and Burstein (2008), intermediate firms consider the impacts of their decisions in the outcomes of their sectors, in which a discrete number of firms compete, but overlook the impacts on aggregate variables over the continuum of sectors, which are considered only by final good producers.

We follow a different approach, assuming each firm $i \in [0, N]$ considers $\partial Y / \partial Y_i = q \in (0, 1]$. Hence, firms pay at least some attention to their impacts on aggregate outcomes. They may be fully aware of them ($q = 1$) but cannot simply ignore them ($q = 0$). We essentially keep all other assumptions from the discrete model, with only minor adjustments. For instance, we again consider the unique equilibrium in which a firm i is active if and only if $A_i \geq \underline{A}$ for some productivity cutoff $\underline{A}$.[11] Moreover, to impose inefficient technologies are common knowledge, we assume the mass of low-productivity firms is finite but sufficiently large to the extent that not all firms can be active simultaneously. Finally, it is worth noting that $\eta > 1$ no longer guarantees the SOC for firms' profit maximization, but it holds if $\eta > 0$ or if $q \in (0, 1]$ is low.[12] In the following, we simply assume that these parameters satisfy that condition.

Under such conditions, all equations presented in Section 2.2.1 for the discrete model would still be valid, with (2.6), (2.7), (2.11), and (2.12) holding *exactly*, if one replaces (i) η by η/q and (ii) sums by integrals (e.g., $Y \equiv \int_0^N Y_i di$ and $\mathrm{E}_a\left(h(A)\right) = \int_{\underline{A}}^{\overline{A}} h(A) \frac{g(A)}{1-G(\underline{A})} dA$, where g is now a density function). They are exactly valid as now $s(\underline{A}) = 0$ holds in equilibrium because marginal adjustments in the number of active firms are allowed in this case.[13] Since the two models have essentially the same equations, they have similar properties. In particular, it is easy to show that (i) $\Omega \in (0, 1]$, (ii) Ω strictly increases with $\underline{A}$, and (iii) $\Omega \to 1$ when $\underline{A} \to \overline{A}$.[14] In short, the continuous model is essentially an exact version of the discrete one.

---

[11] A situation in which not all firms with productivity $\underline{A}$ are active could not occur with a continuum of firms. As a consequence, in this case, we do not need to rule out such situations to obtain this result.

[12] For a discussion on this matter, please refer to Appendix 2.B.6.

[13] In Appendix 2.B.3, we provide a formal proof for why $s(\underline{A}) = 0$ holds in the continuous model equilibrium.

[14] See Appendix 2.B.4 for a proof of such properties.

2.3 Quantification strategy and baseline parameters

In this section, we outline the calibration strategy. Our primary empirical objective is to compute allocative efficiency Ω, which solely requires the distribution of active firms' productivity (Equation (2.11)). Thus, we aim to use real-world data to pin down the distributional parameters. We propose a calibration procedure in which, given a distributional shape, we seek $\underline{A}$ and $\overline{A}$ by matching (i) aggregate TFP $\overline{A}\Omega$ and (ii) cost-weighted average of firm-level markups μ. Owing to the simplicity of our model, we establish necessary and sufficient conditions for this procedure to work properly, achieving an exact match of both target moments. Furthermore, we elucidate the close connection between Ω and μ. Before presenting this calibration strategy, we discuss our distributional assumption and the computation of those moments using real-world data. Afterward, we present our baseline choices for the parameters, including the one that determines the shape of firms' productivity distribution.

2.3.1 Distributional assumption

We consider a continuous distribution of firm productivity, consistent with the continuous model of Section 2.2.2.[15] Consistent with much of the literature (Melitz; Redding, 2015; Edmond; Midrigan; Xu, 2015; Edmond; Midrigan; Xu, 2022), we employ a Pareto distribution with shape parameter $k \neq 0$, truncated within the range $A \in \left[\underline{\underline{A}}, \overline{A}\right]$.[16] It is easy to show that $A \in \left[\underline{A}, \overline{A}\right]$ is also truncated Pareto distributed with shape parameter $k \neq 0$, with density $\tilde{g}(A) \equiv \frac{g(A)}{1-G(\underline{A})} = k\left(\frac{\underline{A}^k \overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1}$.[17] Note this density is strictly increasing in A for $k < -1$, constant for $k = -1$ (Uniform distribution), and strictly decreasing for $k > -1$, $k \neq 0$.

[15] Alternatively, one can interpret that we are using the discrete model of Section 2.2.1 but approximating the expected values of functions of A included in key expressions (e.g., Equation (2.11)) using a continuous distribution as we do not have firm-level data and thus do not know the discrete support.

[16] In the non-truncated Pareto distribution, k should be strictly greater than 0, but in its truncated version k could also be strictly negative.

[17] Consistently, Melitz and Redding (2015, p.24) argue that "a key feature of a Pareto distributed random variable is that it retains the same distribution and shape parameter k whenever it is truncated from below."

Under this density, we show in Proposition 2.D.5 that

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \begin{cases} \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right) & \text{, if } k+j \neq 0 \\ \left(\frac{k\tilde{A}^k}{\tilde{A}^k-1}\right)\ln \tilde{A} & \text{, if } k+j = 0 \end{cases} \tag{2.13}$$

for $k \neq 0$, $j \in \mathbb{N} \setminus \{0\}$, and $\tilde{A} \equiv \overline{A}/\underline{A} > 1$.

2.3.2 Computing the target moments using real-world data

We consider two target moments: (i) the aggregate TFP $\overline{A}\Omega$ and (ii) the cost-weighted average of firm-level markups μ. These moments can be easily obtained using standard macroeconomic data and the parameter α. On the one hand, TFP is backed out as a residual in the aggregate production function (2.8): $\overline{A}\Omega = \frac{Y}{K^\alpha H^{1-\alpha}}$. On the other hand, $\mu = \sum_{i=1}^N \theta_i \mu_i = \sum_{i=1}^N \theta_i p \frac{MPH_i}{w} = \sum_{i=1}^N \theta_i (1-\alpha) \frac{Y_i p}{\theta_i H w} = \frac{1-\alpha}{LS}$, where MPH_i is the marginal product of human capital, $MC_i = \frac{w}{MPH_i}$ as firms use inputs optimally, and $LS \equiv \frac{Hw}{Yp}$ is the labor share of national income. Note $\mu_i = \frac{1-\alpha}{LS_i}$, which is exactly Hall (1988) expression if one considers a Cobb-Douglas production function and human capital as a variable input. Hence, our method to gauge the average markup is closely related to the recent literature that uses Hall (1988) results to estimate firm-level markups (Loecker; Warzynski, 2012; Loecker; Eeckhout, 2018; Loecker; Eeckhout; Unger, 2020; Traina, 2018; Calligaris; Criscuolo; Marcolin, 2018; Autor et al., 2020).[18]

2.3.3 Calibration algorithm

Given our distributional assumption and the computed data moments, we calculate $\underline{A}$ and $\overline{A}$ by matching (i) aggregate TFP $\overline{A}\Omega$ and (ii) cost-weighted average markup μ, in each evaluated period. Hence, similarly to Loecker, Eeckhout and Mongey (2021), even though the model is static, time-varying results can be

[18] Edmond, Midrigan and Xu (2015) and Baqaee and Farhi (2020) use sales-weighted harmonic average instead of cost-weighted arithmetic average. However, these two measures are equivalent here, since $\mu_i = \frac{1-\alpha}{LS_i} \to \frac{s_i}{\mu_i} = \frac{Y_i/Y}{(1-\alpha)/LS_i} = \frac{H_i w/(Yp)}{1-\alpha}$ and thus the harmonic average markup $\left(\sum_{i=1}^N \frac{s_i}{\mu_i}\right)^{-1} = \frac{1-\alpha}{Hw/(Yp)} = \frac{1-\alpha}{LS} = \mu$. Similar results apply to the model of Edmond, Midrigan and Xu (2022), in which these two average markups are also equal, given by an equivalent function of the aggregate labor share.

obtained due to period-by-period estimation. Formally, given $\overline{A}\Omega$ and μ, a solution for $\underline{A}$ and $\overline{A}$, if it exists, can be obtained from the following algorithm:

1. Given Equation (2.13), calculate $\tilde{A} \equiv \overline{A}/\underline{A}$ by solving numerically
$$\mu = \frac{1 - \mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]} \tag{2.14}$$
which is obtained from (2.11) and (2.12) holding exactly as in the continuous model.

2. From Equation (2.12) holding exactly, compute $\underline{A} = \overline{A}\Omega/\mu$.

3. Given $\tilde{A} \equiv \overline{A}/\underline{A}$ and $\underline{A}$ from the previous steps, calculate $\overline{A} = \tilde{A} \times \underline{A}$.

In Appendix 2.D, we establish necessary and sufficient conditions for this algorithm to work properly, achieving an exact match of both target moments. The challenge is to show that a unique solution exists for the first-step problem or, equivalently, that (2.14) implicitly defines $\tilde{A} \equiv \overline{A}/\underline{A}$ as a well-defined function of μ. Referring to Proposition 2.D.8, it becomes evident that a unique solution exists if and only if $\mu > 1$ and $k < 2/(\mu - 1)$.[19]

Note computing $\tilde{A}$ from (2.14) only requires data on average markup μ and thus allocative efficiency Ω does not depend on TFP data either, since (2.11) is just a function of $\tilde{A}$ under (2.13). TFP data are used only to pin down $\underline{A}$ and $\overline{A}$. Therefore, to decompose the TFP, we first estimate Ω using data on $\mu = \frac{1-\alpha}{LS}$, computing then $\overline{A}$ from the observed TFP $\overline{A}\Omega$. As a consequence, in this model, the residual of the production function is not the TFP $\overline{A}\Omega$ itself, but rather only its technology component $\overline{A}$, which is a cleaner residual as it is free of misallocation effects.

Figure 2.1 plots the function Ω of μ for truncated Pareto distributions with $k = 3, 5, 9$ and an Uniform distribution ($k = -1$). Several things are worth noting about it. First, Ω is strictly decreasing in $\mu = \frac{1-\alpha}{LS}$ and thus strictly increasing in

[19] For an arbitrary continuous truncated distribution, a unique solution exists if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$ (Proposition 2.D.4). By employing (2.14), it is possible to compute $\lim_{\overline{A}\to+\infty} \mu$, albeit contingent upon the distributional assumption. For example, for the Pareto case, this is done in Proposition 2.D.7.

LS. In particular, $\Omega \to 1^-$ when $\mu \to 1^+$. Essentially, this function converts one measure of distance from the efficient equilibrium into another. While μ represents the distance of average markup from the competitive, efficient, level $\mu = 1$, allocative efficiency Ω quantifies the deviation of aggregate TFP $\overline{A}\Omega$ from its optimal level, $\overline{A}$. Building on that interpretation, it seems intuitive that a lower average markup $\mu > 1$, indicating a less distorted economy, would be associated with enhanced allocative efficiency Ω. Second, given a time series of μ, a lower k would imply a higher and less volatile estimated Ω.

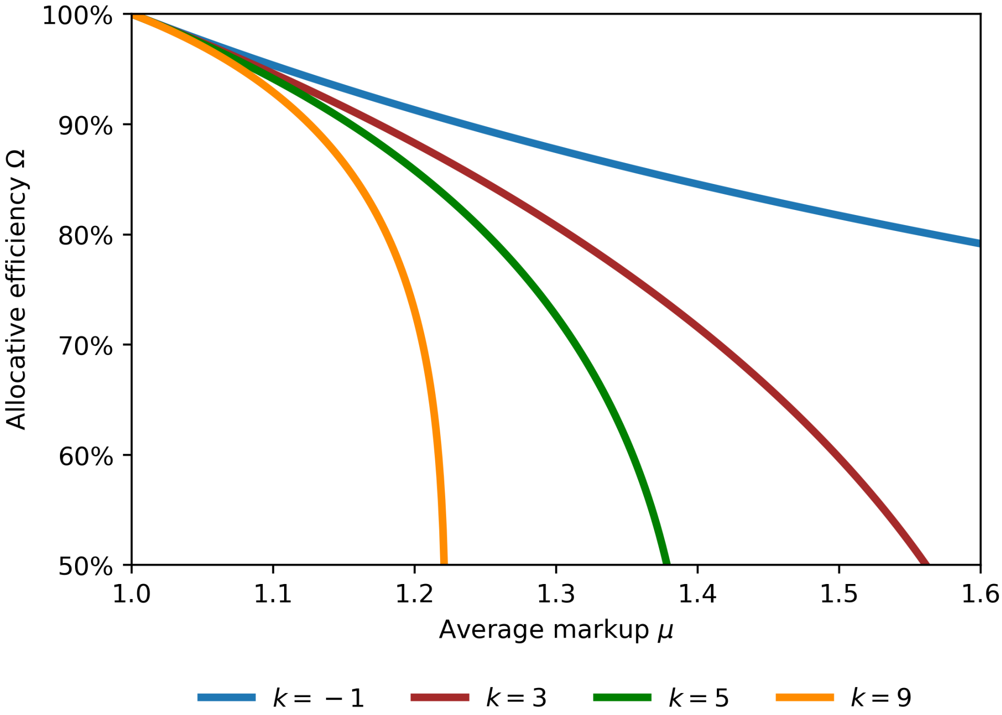


Figure 2.1 – Allocative efficiency vs. average markup.

2.3.4 Baseline parameters

In our baseline calibration, we set $k = -1$, that is, we suppose firm productivity is uniformly distributed. Note, in this case, the calibration algorithm will properly work if and only if $\mu > 1$, since $k < \frac{2}{\mu-1}$ will always hold. This choice of k is justified for two reasons.

First, $k = -1$ yields the most conservative reasonable TFP decomposition compared to standard accounting exercises, in which all TFP variation is attributed to technology. To see that, note we obtain a TFP decomposition that is closer to standard ones by choosing a lower k, as Ω would increase towards one and become increasingly stable over time (Figure 2.1). Indeed, from Proposition 2.D.9, $\Omega \to 1$ when $k \to -\infty$ and thus our model's residual $\overline{A}\Omega$ converges to the standard accounting residual $\overline{A}$ in this limit case. Given that, it would be useful to find a lower bound for k. Supposing high-productivity firms are relatively scarce, this lower bound is exactly $k = -1$. After all, as discussed in Section 2.3.1, the truncated Pareto density is downward sloping only for $k > -1$, $k \neq 0$, being upward sloping for $k < -1$ and constant for $k = -1$ (Uniform distribution). Besides being economically reasonable, this assumption is commonly adopted, often by imposing a non-truncated Pareto distribution for firms' productivity.

Second, we show in Appendix 2.E that, under $k = -1$, the estimates of allocative efficiency *growth* are highly robust to (i) the level of LS and (ii) the choice of α, as soon as $\mu = \frac{1-\alpha}{LS} > 1$. These are interesting features since (i) estimating α is not an easy task and (ii) gauging the level of LS is not straightforward, particularly for developing countries, given the difficulty of identifying the labor share of self-employment income (Gollin, 2002).

Naturally, this does not mean that the choice of α is irrelevant under $k = -1$. Besides possibly changing the number of country-year observations that fulfill the model's necessary condition $\mu = \frac{1-\alpha}{LS} > 1$, α is required to pin down (i) the TFP $\overline{A}\Omega$, (ii) the (level of the) allocative efficiency Ω, and (iii) the technology frontier $\overline{A}$. As a result, the choice of α is fundamental for decomposing the TFP $\overline{A}\Omega$ growth as it alters the TFP $\overline{A}\Omega$ growth itself and consequently the growth of the residual $\overline{A}$. In short, even for $k = -1$, one should choose α. Our baseline calibration employs the standard $\alpha = 1/3$ for all countries and years.

2.4 Growth in nations at the frontier

In this section, we examine economic growth in the United States. Following Jones (2016), we take it as a proxy for the performance of nations situated at the world income frontier.

2.4.1 Data and parameters

We use annual data from the Penn World Table 10.01 (Feenstra; Inklaar; Timmer, 2015) to measure the US variables. Output Y is the real GDP at constant national prices (*rgdpna* variable), and physical capital K is the capital services at constant national prices (*rkna* variable). We measure labor L as the aggregate amount of time worked, which is the product of two variables: (i) the number of persons engaged (*emp* variable) and (ii) the average annual hours worked by persons engaged (*avh* variable). Multiplying L by the human capital index (*hc* variable), based on years of schooling and returns to education, we obtain the stock of human capital H. Finally, LS is the share of labor compensation in GDP at current national prices (*labsh* variable). For the US, it is always equal to the preferred labor share measure of the Penn World Table 10.01, which assumes self-employed workers use labor and physical capital in the same proportion as the rest of the economy (Feenstra; Inklaar; Timmer, 2015). The data for these variables span from 1950 to 2019, except for the physical capital stock, available only from 1954 onward.

As discussed in Section 2.3.4, we use the standard $\alpha = 1/3$ in the baseline calibration. Alternatively, we could calibrate it from data. Since firms use inputs optimally, α equals the cost share of capital, that is, $\alpha = \frac{Kr}{Kr+Hw}$.[20] Using factor income data for the US nonfinancial corporate sector from Barkai (2020), we estimate $\alpha = 0.31$.[21] As it is only slightly lower than the standard calibration, we maintain $\alpha = 1/3$ throughout the main body of the paper. In Appendix 2.G, we discuss this alternative calibration procedure in detail, demonstrating that our main conclusions are robust to this lower α.

Given the labor share LS and the parameter α, we compute the average markup $\mu = \frac{1-\alpha}{LS}$. Figure 2.2 plots the results from 1950 to 2019, showing an increase

[20] For a formal proof, please refer to Appendix 2.A.4.

[21] Similarly, Loecker, Eeckhout and Unger (2020) and Edmond, Midrigan and Xu (2022) use cost data to gauge sectoral elasticity of output with respect to labor, which is equivalent to $1 - \alpha$ in our Cobb-Douglas case.

in the average markup. This is consistent with evidence supporting higher firms' markup power in the US (Loecker; Eeckhout; Unger, 2020; Baqaee; Farhi, 2020; Loecker; Eeckhout, 2018).

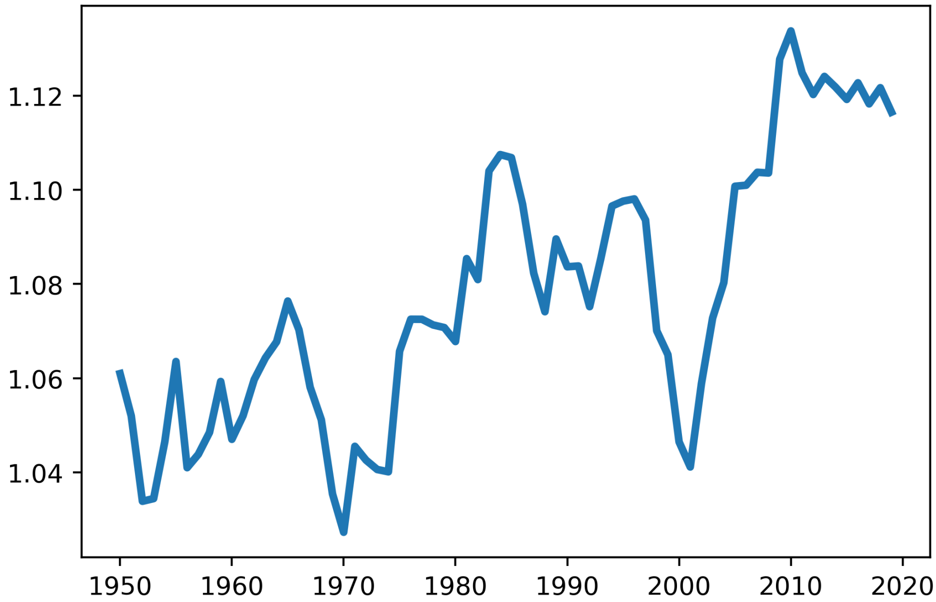


Figure 2.2 – Average markup in the United States.

In the baseline calibration, we use $k = -1$ due to the desirable features presented in Section 2.3.4. However, one may be interested in evaluating the reasonableness of this assumption for the United States. Edmond, Midrigan and Xu (2022) estimate misallocation using an oligopoly model calibrated for the US. Rather than committing to a single average markup μ, they calibrate it for μ equal to 1.05, 1.15, 1.25, or 1.35. Table 2.1 compares the value-added aggregate productivity losses shown in their Table 6 with our own estimates for $k = -1, 3, 5$, computed accordingly as $-(\overline{A}\Omega - \overline{A})/\overline{A} = (1 - \Omega)$. As can be seen, our estimates are closer to theirs for $k = -1$, particularly for $\mu = 1.15$ when the results are practically the same. This is a very significant result considering that their model is primarily calibrated using recent data, a timeframe during which $\mu = \frac{1-\alpha}{LS}$ consistently hovers around 1.15 (Figure 2.2). For instance, to gauge productivity dispersion, a crucial factor in estimating misallocation, they target measures of concentration in 2012, when we find $\mu = 1.12$. This suggests $k = -1$ is indeed a reasonable assumption for the US, which serves as an important validation of our model, particularly considering that the oligopoly model of Edmond, Midrigan and Xu (2022) is more flexible, general, requiring firm-level data for calibration. In any case, as $k = -1$ is an economic lower bound for k in our model, we choose to test also the higher values shown in Table 2.1.

Table 2.1 – Aggregate productivity losses, %

		Average markup μ			
		1.05	1.15	1.25	1.35
Edmond, Midrigan and Xu (2022) estimates (Oligopoly model)		5.55	6.85	8.89	12.52
Own estimates	$k = -1$	2.41	6.75	10.54	13.90
	$k = 3$	2.60	8.43	15.31	23.58
	$k = 5$	2.70	9.62	19.87	38.77

2.4.2 Allocative efficiency

Figure 2.3 plots the estimated allocative efficiency Ω for each k. Several things are worth noting about these estimates. First, since Ω is just a transformation of μ, the estimated series are highly correlated with each other, showing similar trends.[22] In particular, all estimates suggest the allocative efficiency Ω is decreasing over time, especially after 2000, reflecting the lower labor share LS and thus the higher average markup $\mu = \frac{1-\alpha}{LS}$ in recent years (Figure 2.2). This is consistent with Baqaee and Farhi (2020) that finds an increase in the distance from the optimal allocation between 1997 and 2015 in the US. Second, consistent with Figure 2.1, Ω is higher and more stable for lower k. In any case, the estimates are relatively similar, especially considering that $k = 5$ is probably very extreme given the results of Table 2.1.

[22] This is especially clear at the beginning of the sample, because the estimate of Ω is less sensitive to k for a lower markup μ (Figure 2.1), as observed in sample's initial years (Figure 2.2).

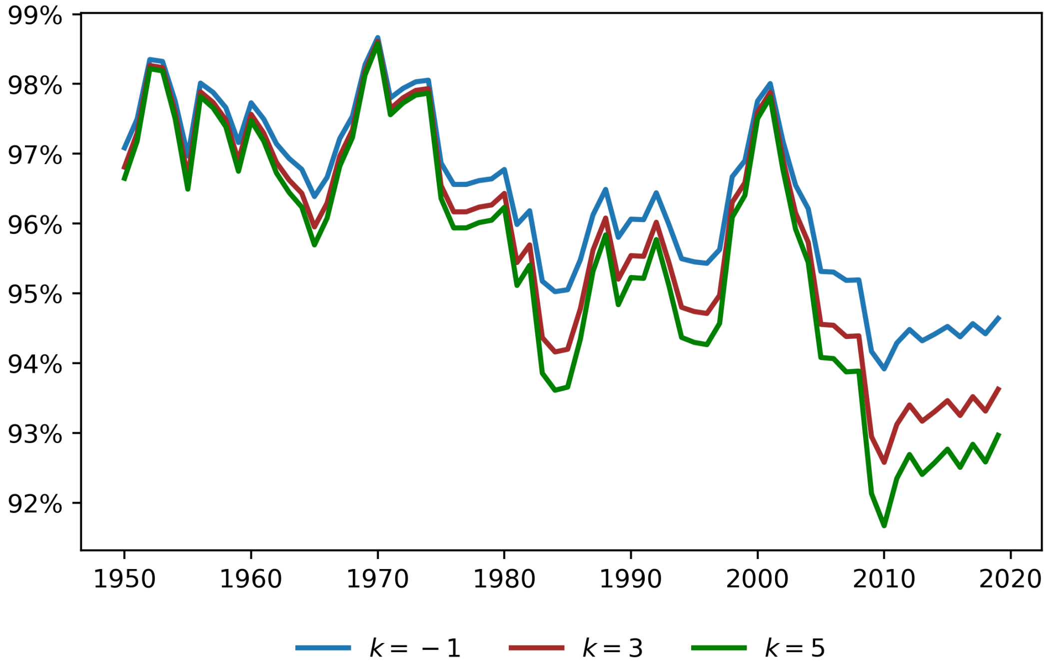


Figure 2.3 – Allocative efficiency in the United States.

2.4.3 Growth accounting

To understand the sources of economic growth, one can use directly the aggregate production function (2.8). However, in this case, there would be no clear separation between the components since the accumulation of physical capital is typically affected by productivity (e.g., consider a standard Solow model). A better alternative followed by Jones (2016) is to use

$$\frac{Y_t}{L_t} = \left(\frac{K_t}{Y_t}\right)^{\alpha\beta} \left(\frac{H_t}{L_t}\right) \left(\overline{A}_t\Omega_t\right)^{\beta} \tag{2.15}$$

which can be easily obtained from (2.8) if one adds time subscripts and defines $\beta \equiv \frac{1}{1-\alpha}$. In this form, output per hour Y_t/L_t grows due to the growth of (i) physical capital-output ratio through $(K_t/Y_t)^{\alpha\beta}$, (ii) human capital per hour H_t/L_t, and (iii) labor-augmenting TFP $\left(\overline{A}_t\Omega_t\right)^{\beta}$. Jones (2016) points out that the contributions from productivity and physical capital are separated in (2.15) in a way that they were not in (2.8). After all, as suggested by a neoclassical growth model, the physical capital-output ratio does not depend on TFP, at least in the long run.

Table 2.2 presents the growth accounting exercise based on (2.15). Essentially, we repeat the exercise shown in Table 3 of Jones (2016), but bringing the TFP decomposition to the analysis, finding qualitatively similar results.[23] First of all, most of the long-run growth in the output per hour comes from the TFP, although Jones (2016) finds an even higher contribution from it. Human capital is also relevant, while the physical capital-output ratio contributes little to output growth. Moreover, we also find faster output and TFP growth until 1973, although in our decomposition some of this output acceleration was caused by human capital. According to Gordon (2000), this period of outstanding TFP growth can be dated back to 1913, showcasing the impact of the great inventions of the late nineteenth and early twentieth centuries. These innovations notably encompass electricity and the internal combustion engine, but also include chemicals, movies, radio, and indoor plumbing. This boom period is followed by slower economic growth between 1973 and 1995, mainly due to a productivity slowdown. Between 1995 and 2007, economic growth accelerated again, driven by productivity growth, "[...] coinciding with the dot-com boom and the rise in the importance of information technology" (Jones, 2016, p.11). Finally, we also find lackluster economic growth from 2007 onward, in the wake of the Great Recession and among slower TFP growth. Fernald (2015) presents compelling evidence that this productivity slowdown began before the global financial crisis, tracing back to 2003, reflecting "[...] a retreat from the exceptional, but temporary, information technology-fueled pace from the mid-1990s to early in the twenty-first century" (Fernald, 2015, p.1). Similar results are found in Byrne, Oliner and Sichel (2013).

All in all, TFP seems to be the key variable in explaining economic growth in both the short and long run. The problem is that the TFP is usually a residual, "a measure of our ignorance." This is not true here, where the residual is not the TFP $\overline{A}\Omega$ itself but only the technology component $\overline{A}$. Allocative efficiency Ω is gauged from the average markup $\mu = \frac{1-\alpha}{LS}$. Looking at Table 2.2, we see that long-run TFP growth is driven almost solely by its technology component. Indeed, between 1954 and 2019, labor-augmenting TFP grew 1.1% per year, almost equal

[23] This occurs despite the distinct data sources used. Jones (2016) uses 1948-2013 data for the US private business sector from the Bureau of Labor Statistics, while we use 1954-2019 US national data from the Penn World Table 10.01.

Table 2.2 – Growth accounting for the United States

Period	$\frac{Y}{L}$	Y/L components			Labor-aug. TFP $\left(\overline{A}\Omega\right)^{\beta}$ components					
					$k=-1$		$k=3$		$k=5$	
		$\left(\frac{K}{Y}\right)^{\alpha\beta}$	$\frac{H}{L}$	$\left(\overline{A}\Omega\right)^{\beta}$	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}
1954-2019	1.9	0.2	0.5	1.1	1.2	-0.1	1.2	-0.1	1.2	-0.1
1954-2013	1.9	0.2	0.6	1.2	1.2	-0.1	1.3	-0.1	1.3	-0.1
1954-1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0	1.7	0.0
1973-1990	1.3	0.3	0.5	0.5	0.7	-0.2	0.7	-0.2	0.8	-0.2
1990-1995	1.6	0.2	0.5	0.9	1.1	-0.2	1.2	-0.3	1.2	-0.3
1995-2000	2.2	0.2	0.3	1.7	1.0	0.7	0.8	0.9	0.7	1.0
2000-2007	2.2	0.4	0.3	1.5	2.0	-0.6	2.2	-0.7	2.3	-0.8
2007-2013	1.3	0.5	0.3	0.5	0.7	-0.2	0.8	-0.3	0.9	-0.4
2013-2019	1.0	-0.2	0.1	1.0	0.9	0.1	0.9	0.1	0.8	0.2

Note: Logarithmic approximation of average annual growth rates (in percent).

to the 1.2% annual growth in the technology component. Changes in allocative efficiency represented just a small drag on growth (-0.1% per year).

However, misallocation can be much more relevant for shorter periods of time. For instance, during 1995-2000 and 2000-2007, amid the IT boom, TFP grew at similar high rates, but for very distinct reasons. TFP growth between 1995 and 2000 is attributable almost equally to both of its components, whereas from 2000 to 2007, it was solely driven by technology, with misallocation worsening throughout the period. These findings appear to align with sector-level TFP estimates. Fernald (2015) identifies substantial productivity gains in IT-producing industries during 1995-2000, while noting only modest improvements in IT-intensive sectors. Similarly, Gordon (2000) demonstrates that productivity enhancements from 1995 to 1999 were primarily concentrated in computer and computer-related semiconductor manufacturing. As a result, given that IT-producing sectors accounted for only 4.9% of private business value-added between 1988 and 2011 (Fernald, 2015), one should not anticipate significant IT-induced aggregate productivity gains in the late 1990s. While the behavior of aggregate TFP does not align with this reasoning, our decomposition analysis does. Specifically, technology improved during 1995-2000 at essentially the same pace observed between 1973 and 1995. The TFP growth from 1995 to 2000 appears to originate from elsewhere, reflecting better diffusion of technology across firms and consequently improved resource allocation. We find significant technology improvements only after 2000, which is consistent with Fernald (2015) showing that TFP of the *large* IT-intensive industries surged in the early twenty-first century.[24] Nevertheless, the decline in allocative efficiency during 2000-2007 suggests that these IT-induced technology improvements were not effectively disseminated across all firms.

Another example is given by the comparison between 2007-2013 and 2013-2019, when the TFP annual growth doubled, from 0.5% to 1%, but the technology component grew much more similarly across the periods due to the behavior of misallocation. Indeed, this fact seems to hold more generally, as the growth of the technology component is more stable, around 1% per year. The main exceptions are the periods of 1954-1973 and 2000-2007 when technology contributed around 2% per year. This is consistent with the argument of Fernald (2015) that the exceptional growth during these periods is what appears unusual and warrants investigation. The subsequent slowdowns are merely the flip side of these speedups, representing the return to a normal growth pace. Additionally, among these unusual periods, it is worth noting that we find more rapid technology improvements from 2000 to 2007 amidst developments in IT, even though TFP grew faster between 1954 and 1973.

2.5 Economic performance around the world

In Section 2.4, we examined the growth of nations situated at the global income frontier, employing the United States as a proxy. Now, our focus shifts to assessing economic performance around the world, which we approach from two distinct perspectives. First, we conduct a development accounting exercise to understand the differences in income across countries. Second, we reexamine the diffusion of growth across nations, actively seeking evidence of convergence not only

[24] According to Fernald (2015), during the 1988-2011 period, these IT-intensive sectors accounted for 34.9% of private business value-added.

in income but also in TFP. In addition to updating previous similar assessments, our exercises yield fresh insights and discussions by incorporating misallocation estimates.

2.5.1 Data and parameters

As for frontier growth, variables required by the model are taken from the Penn World Table 10.01, which contains annual data between 1950 and 2019 for 183 countries. While development accounting solely requires variables' levels, assessing convergence involves examining the cross-country correlation between a variable's growth over a certain period and its level in the initial year. Consequently, it is essential to choose appropriate measures for both level and growth analyses, which may differ.

For level analyses, we consider two distinct data sets: one with the most accurate proxies but fewer country-year observations and another with broader coverage, albeit relying on worse measures of the variables of interest. In the data set with the greatest sample coverage, Y is the output-side real GDP at current PPPs (*cgdpo* variable), K is the capital stock at current PPPs (*cn* variable), L is the number of persons engaged (*emp* variable), H is L multiplied by the human capital index (*hc* variable), and LS is the share of labor compensation in GDP at current national prices (*labsh* variable). In the data set that uses the best proxies, Y is measured in the same way, while K is the capital services levels at current PPPs (*ck* variable). To obtain L and H, we multiply the previous proxies by the average annual hours worked by persons engaged (*avh* variable). Finally, LS is the share of labor compensation in GDP at current national prices as computed using their adjustment 2 method (*lab_sh2* variable, available at the labor detail database). This is the preferred labor share measure of the Penn World Table 10.01, in which self-employed workers are assumed to use labor and physical capital in the same proportion as the rest of the economy (Feenstra; Inklaar; Timmer, 2015).

For growth analyses, we measure Y as the real GDP at constant national prices (*rgdpna* variable) and K as the capital stock at constant national prices (*rnna* variable). L and H are measured in the same way as in the data set for level analyses with the greatest sample coverage. To pin down allocative efficiency improvements, we need to measure the changes in labor share properly. However, to ensure complete coverage over the years, Feenstra, Inklaar and Timmer (2015) make some interpolations and extrapolations that distort the labor share growth as they (i) assume labor shares remain constant or (ii) linearly interpolate if there are missing years in the middle of the sample. To avoid such distortions, we choose to discard any imputed data, which are identified in their labor detail database. These imputations are done for each of the following labor share measures, which differ in the way they treat the income of self-employed workers (also known as mixed income):

1. Naïve share: $LS = \frac{\text{labor compensation}}{\text{GDP at basic prices}}$
2. Adjustment 1, mixed income: $LS = \frac{\text{labor compensation+mixed income}}{\text{GDP at basic prices}}$
3. Adjustment 2, part mixed income: $LS = \frac{\text{labor compensation}}{\text{GDP at basic prices}-\text{mixed income}}$
4. Adjustment 3, average wage: $LS = \frac{\left(\frac{\text{labor compensation}}{\text{\# employees}}\right)(\text{\# employees and self-employed})}{\text{GDP at basic prices}}$
5. Adjustment 4, agriculture: $LS = \frac{\text{labor compensation+value added in agriculture}}{\text{GDP at basic prices}}$

Subsequently, they apply a set of rules to select a measure for each country and year, obtaining their final estimate (*labsh* variable). Hence, the selected measure may vary from period to period, which could also distort the labor share growth. Given that, we do not use the *labsh* variable for growth analyses but choose among the above measures (without imputations). As argued for the development accounting case, adjustment 2 share (*lab_sh2* variable) is the best proxy available. However, it requires mixed income data available for only 79 countries. Consequently, we choose to measure LS using the naïve share (*comp_sh* variable, from the labor detail database) since it is available for 139 countries and seems to serve as a good proxy for the labor share growth.

This last fact is illustrated in Figure 2.4, which plots labor share annual growth $\Delta \ln LS$ for adjustment 2 method against each of the other four measures (all without imputations), considering only country-year observations in which all five labor share growth measures are available. We also plot the 45-degree line and the Ordinary Least Squares (OLS) linear trend, whose associated centered

R^2 is shown inside each plot. Since the adjustment 1 measure also requires mixed income data, it cannot be a solution to the small sample problem of the preferred adjustment 2 share. Among the other alternatives, the naïve share has the best linear fit, with its linear trend being essentially equal to the 45-degree line. Hence, the naïve share provides the best proxy for the growth in the preferred adjustment 2 measure, suggesting it can serve as a good proxy for the labor share growth. However, the naïve share is known to underestimate the labor share as it does not allocate any income from self-employed workers to labor (Gollin, 2002). As a consequence, this labor share would overestimate the average markup $\mu = \frac{1-\alpha}{LS}$, which can be relevant even for growth analyses. For instance, for $k > 0$, a country may even be dropped from the sample since the necessary condition $k < \frac{2}{\mu-1}$ could not hold. To address this issue, we multiply the naïve share (without imputation) by a country-specific constant that aligns its average with the mean of the *labsh* variable (with imputation), using the longest coincident period available. By doing that, we maintain the naïve share growth but correct its level.

We constructed this data set for growth analyses focusing on maximizing sample coverage. Additionally, we considered another data set that employs the best proxies available for growth analyses.[25] However, it is impractical to evaluate convergence using the best proxies for both level and growth analyses due to the significantly reduced sample size. As a result, we present convergence results only for the largest data sets.

We finish the data description by commenting on two variables used in both level and growth analyses. First, GDP per person, which is Y divided by the population (*pop* variable from the Penn World Table 10.01). Second, our analyses are focused on non-oil countries. We classify a country as an oil producer in a given year if its oil rents account for more than 10% of GDP in that year. Oil rents as a share of GDP are sourced from the World Development Indicators provided by the World Bank.[26]

[25] In this case, K is the capital services at constant national prices (*rkna* variable), L and H are the best proxies for level analyses, and LS is the adjustment 2 share (*lab_sh2* variable), without imputations.

[26] If oil rents are unavailable for a country at a given year, we utilize data from the most recent non-missing year for that country. If such data is unavailable, we resort to data from the subsequent non-missing year. If this is also unavailable, we assume oil rents to be null.

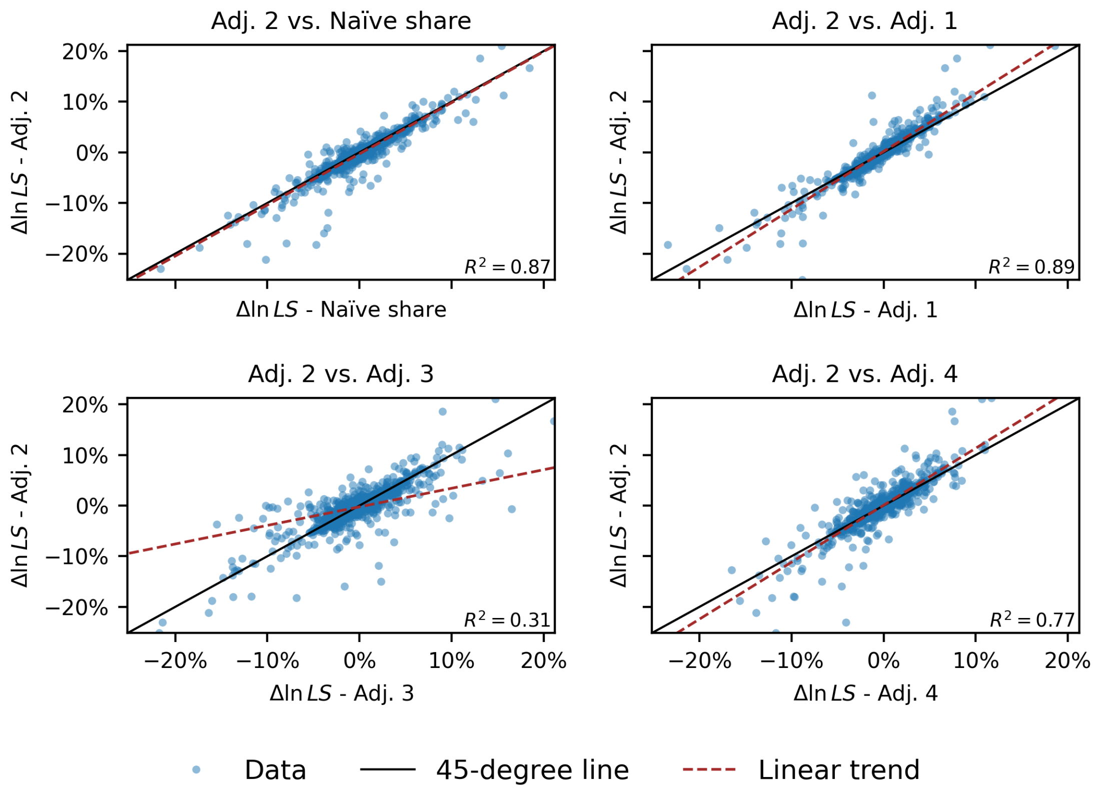


Figure 2.4 – Comparing measures of labor share annual growth.

Regarding the parameters, we use the baseline values $\alpha = 1/3$ and $k = -1$ discussed in Section 2.3.4. Since $k = -1$ is an economic lower bound for k, we also test $k = 3$. It is worth mentioning that we could have used $\alpha = 0.31$ as estimated in Section 2.4.1 using US data from Barkai (2020). However, since it is not so different from $\alpha = 1/3$, we opt to keep to the standard practice employed in development accounting exercises (Caselli, 2005; Caselli, 2016; Jones, 2016).

Figure 2.5 plots the number of non-oil countries for which we could decompose the TFP for each data set, k, and year between 1950 and 2019. Several facts of this graph stand out. First, as expected, the best proxies are available for fewer countries. In particular, the sample coverage of the data set with the best proxies for growth analyses is notably restricted, especially before the mid-1990s. Second, the sample coverage is smaller for $k = 3$ than for $k = -1$, since the necessary condition

$k < \frac{2}{\mu-1}$ always holds under $\mu > 1$ for $k < 0$, but not necessarily for $k > 0$.[27] Third, the sample coverage typically increases over time. The main exception occurs in growth analyses for 2019 when the labor share measures are available only for the US. This does not happen in the data sets for level analyses due to the use of extrapolated data. Fourth, this use of interpolated/extrapolated labor share data in level analyses increases its sample coverage relative to growth analyses, which is especially clear closer to the beginning of the sample.

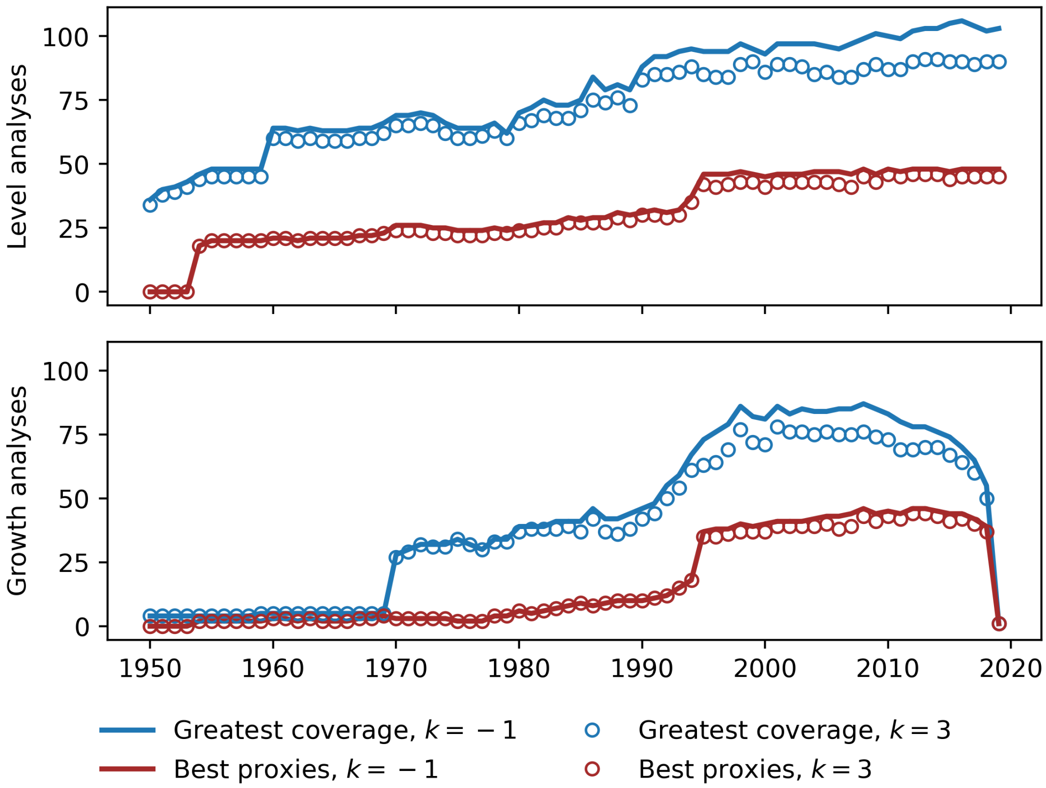


Figure 2.5 – Number of non-oil countries whose TFP could be decomposed.

In the subsequent sections, we leverage these data, employing the TFP decomposition to reexamine key aspects of economic growth. For an initial exploration of the data sets for level analyses, please refer to Appendix 2.H. This appendix features scatter plots depicting the labor-augmenting TFP and its components against GDP per unit of labor for 1965, 1975, 1985, 1995, 2005, and 2015. Appendix

[27] Indeed, for $\mu > 1$ and $k = 3 > 0$, $k < \frac{2}{\mu-1} \leftrightarrow \mu < \frac{2+k}{k} \approx 1.67$.

2.I provides a similar exploration of the data sets for growth analyses, but, in this case, we plot each variable's five-year annual average growth instead of their levels.

2.5.2 Development accounting

Growth accounting aims to identify the proximate determinants of economic growth. It is commonly utilized to evaluate a country's performance over time, as we do for the US in Section 2.4.3. Development accounting applies the same logic but to explain income differences across countries at a specific point in time, typically a year. As Jones (2016) points out, the physical capital-output ratio is remarkably stable across countries. Human capital varies more as poorer countries usually present lower levels of educational attainment, but still modestly so in light of the substantial cross-country income differences. A consequence of these two facts is that wealthier nations exhibit significantly higher productivity, leading to the usual conclusion that income differences are predominantly attributed to TFP (Caselli, 2005; Caselli, 2016; Jones, 2016). Since TFP is typically a residual, "a measure of our ignorance," this suggests that most income variability remains unexplained. However, as our model pins down allocative efficiency from the average markup, the residual becomes only its technology component. This prompts the question: can misallocation offer insights into cross-country income differences? This is the inquiry that we aim to explore in this section.

Following the discussion of Section 2.4.3, our development accounting exercise relies on production function (2.15) instead of (2.8), since, in that case, the contributions from productivity and physical capital are better separated than in this one.[28] From (2.15),

$$\text{Var}(y) = \text{Var}(e) + \text{Var}(u) + 2\text{Cov}(e, u) \tag{2.16}$$

where $\text{Var}(\cdot)$ and $\text{Cov}(\cdot)$ are, respectively, sample variance and covariance across countries at a given year, while $y \equiv \ln(Y/L)$ is the natural logarithm of the income per unit of labor, e represents the explained portion of y, and u denotes the unexplained part of y. We initially consider the standard case where TFP is a residual, such that $e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(H/L)\right]$ and $u \equiv \ln\left[\left(\overline{A}\Omega\right)^{\frac{1}{1-\alpha}}\right]$. Based on this

[28] However, here one should interpret the subscript t in (2.15) as denoting country, not time.

variance decomposition, we can identify how successful the factors of production are in explaining cross-country income differences. We use two standard measures from the development accounting literature (Klenow; Rodríguez-Clare, 1997; Caselli, 2005):

$$success_1 \equiv \frac{\mathrm{Var}(e)}{\mathrm{Var}(y)} \tag{2.17}$$

$$success_2 \equiv \frac{\mathrm{Var}(e) + \mathrm{Cov}(e,u)}{\mathrm{Var}(y)} = \frac{\mathrm{Cov}(e,y)}{\mathrm{Var}(y)} \tag{2.18}$$

Undoubtedly, the variance term associated with the factors of production should be incorporated into the numerator of such success measures. However, the appropriate treatment of the covariance term is less clear and distinguishes these two metrics. In the first measure, the entire covariance term is assigned to TFP, whereas the second measure distributes this term equally between factors of production and TFP. Note that this second measure can also be computed as the OLS slope coefficient from regressing the factors term e on the output per unit of labor y. Analogous reasoning applies to its complement, the TFP share. Hence, as Klenow and Rodríguez-Clare (1997, p.80) point out, this variance decomposition amounts to asking: when we see 1% higher output per labor in one country, how much higher we expect factors and TFP terms to be?

Alternatively, we calculate these two success measures with allocative efficiency included in the explained portion of y, consistent with our oligopoly model in which the residual is just the technology component $\overline{A}$ rather than the entire TFP $\overline{A}\Omega$. Formally, we again employ (2.17) and (2.18), but, in this case, using $e \equiv \ln\left[(K/Y)^{\frac{\alpha}{1-\alpha}}(H/L)\,\Omega^{\frac{1}{1-\alpha}}\right]$ and $u \equiv \ln\left[\overline{A}^{\frac{1}{1-\alpha}}\right]$. For both measures, high values indicate that e successfully explains cross-country income differences, while low values denote failure as u is the key variable behind the observed results. Consequently, by computing these metrics both with and without allocative efficiency Ω included in the explained portion e, we can check if misallocation enhances our understanding of cross-country income differences. If the measures are higher when Ω is included in e, we conclude that it enhances our understanding; if they are not, we conclude that it does not.

The results are displayed in Figure 2.6. We use all non-oil countries available for each data set for level analyses, k, and year. Therefore, as previously seen in Figure 2.5, the evaluated countries may change over time in a given plot and across plots for a given year, requiring caution when performing such comparisons. Our primary focus is comparing the measures within a given data set, k and year, when the same sample of countries is used. As can be seen, both measures of success increase when misallocation is included in the explained portion, across all the evaluated years. This is evident in Figure 2.7, which plots the difference between the metrics with and without misallocation. The gains are more pronounced for the first measure. This means that the inclusion of misallocation elevates the variance of the explained portion, thereby enhancing the first metric. However, the covariance between explained and unexplained components diminishes, resulting in smaller improvements in the second measure. Or, equivalently, both the variance of the explained portion and its covariance with income per unit of labor increase, but the former sees a greater rise. Note that this higher covariance implies a positive relationship between allocative efficiency and income, which is illustrated in the scatter plots of Appendix 2.H (Figures 2.H.9, 2.H.10, 2.H.11, and 2.H.12).

It is also noteworthy that the gains are more modest for $k = -1$, typically below 10pp, but it becomes more relevant as k increases. Moreover, we find stronger enhancements when using the best proxies available, with the gains typically doubling in this smaller data set. However, this last result may be reflecting differences in countries' sample rather than the measures themselves, as these increased gains essentially disappear when we consider the same sample of countries in each data set (Figures 2.J.1 and 2.J.2 in Appendix 2.J).[29] In contrast, the higher gains for $k = 3$ are not attributable to sample issues; they instead reflect the characteristics of the model shown in Figure 2.1, with a higher k leading to allocative efficiency estimates that vary more across countries.

In summary, these metrics consistently indicate that misallocation plays a significant role in explaining cross-country income differences, particularly for higher values of k. Despite its significance, the measures of success remain relatively low, typically below 50%, at least for the most recent two decades, which have

[29] In Appendix 2.J, we also depict the success measures from 1990 onwards but under time-invariant countries' samples (Figures 2.J.3 and 2.J.4). Overall, the results exhibit consistency with those presented here.

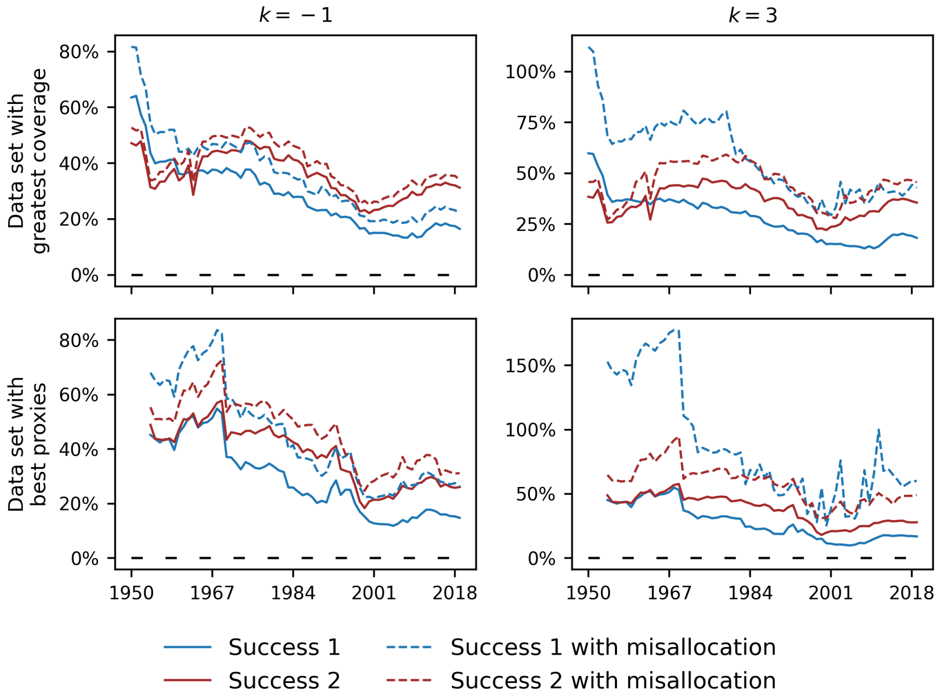


Figure 2.6 – Measures of success, all non-oil countries available.

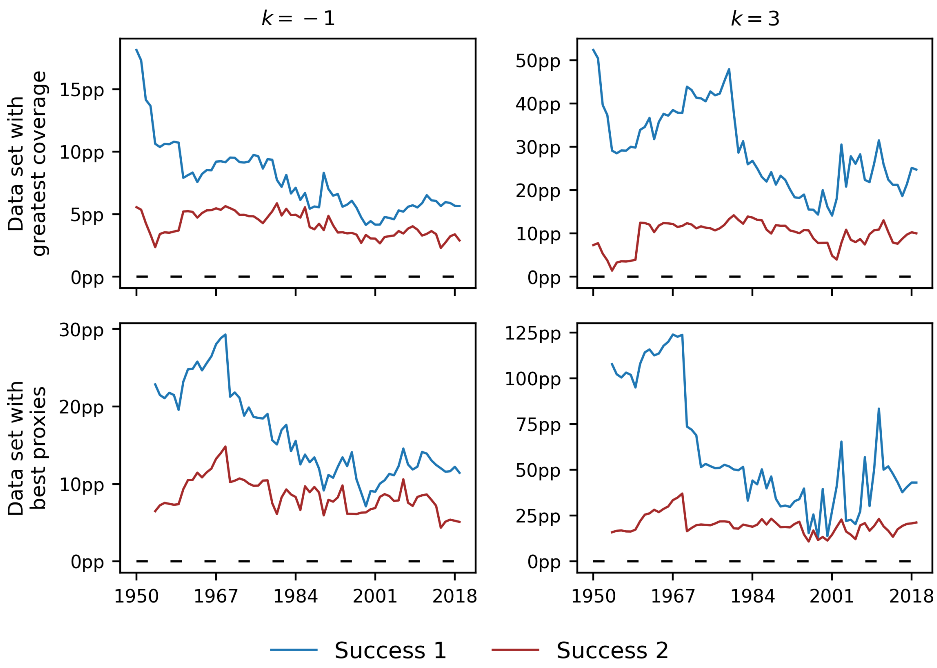


Figure 2.7 – Success gains due to misallocation, all non-oil countries available.

broader sample coverage. In other words, the unexplained portion continues to be substantial, still constituting the majority of observed variability in most cases. Caselli (2005, p.681) states that a "sentence commonly used to summarize the existing literature sounds something like 'differences in efficiency account for at least 50% of differences in per capita income"'. Our results indicate that while market-power-driven misallocation is essential to understanding cross-country income differences, it does not fundamentally alter this statement.

2.5.3 Convergence assessment

In his Figure 26, Jones (2016) evaluates income convergence by comparing the 1960-2011 economic growth rate with GDP per person in 1960 across countries. As these variables do not exhibit a negative correlation, Jones (2016, p.36) concludes "that a simplistic view of convergence does not hold for the world as a whole." In Figure 2.8, we conduct a similar exercise but for 1975-1995, 1975-2005, 1975-2015, 1985-2005, 1985-2015, and 1995-2015. GDP per person growth (level) is from the data set for growth analyses (level analyses) with the greatest coverage. We consider only countries whose TFP could be decomposed in both level and growth data sets under $k = -1$. Note we distinguish between oil-producing and non-oil-producing countries in the plot.[30] For each period, we also show OLS linear trends for non-oil countries and the complete set of countries, whose slope coefficients' estimates and statistical significance are displayed inside each plot. Consistent with Jones (2016), we do not find strong support for income convergence. Indeed, the variables seem to be negatively correlated only for 1995-2015.

In Section 2.5.2, we saw that TFP is key in understanding cross-country output differences, suggesting that the above findings for income may reflect a lack of convergence in TFP. We investigate it in Figures 2.9, 2.10, and 2.11, where we perform convergence assessments for TFP $\overline{A}\Omega$, technology frontier $\overline{A}$, and allocative efficiency Ω, respectively. Similarly to the income case, the convergence hypothesis does not seem to hold for the TFP and both its components except perhaps for 1995-2015, when poorer countries grew faster among swifter technology

[30] In this case, an oil country is defined by having oil rents that account for more than 10% of its GDP in *any* year within the specified period.

advancements and more pronounced gains in allocative efficiency.

A more comprehensive convergence assessment is presented in Figure 2.12, which shows the same non-oil slope estimates displayed in the scatter plots but for each 20-year period commencing in 1970. We additionally depict the 90% confidence interval. All in all, this evidence reaffirms our earlier observations that convergence appears elusive, except possibly in the most recent decades. It is noteworthy that these conclusions remain valid for $k = 3$ (Figure 2.13).[31]

In short, the empirical evidence does not strongly support the convergence hypothesis in either income or TFP, with the lack of convergence being evident in both TFP components. Consequently, countries do not appear to be converging over time to a common degree of allocative efficiency, indicating that the level of efficiency is country-specific even in the long run. Interestingly, this suggests that market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

2.6 Model extensions

We discuss two model extensions in Appendix 2.F. First, we go beyond the Cobb-Douglas production function (2.2) and show that the model's key equations are still valid for an arbitrary well-behaved production function with M factors of production, provided it exhibits (i) constant returns to scale and (ii) Hicks-neutral productivity shifter. Consequently, given data on the TFP $\overline{A}\Omega$ and the average markup μ, we can quantify the model by following precisely the empirical strategy of Section 2.3.3, relying on the same conditions for the existence of a solution for the calibration algorithm. The differences between these models appear only in the computation of the TFP $\overline{A}\Omega$ and the average markup μ from data since (i) the expressions used to compute these two data moments are different and (ii) the calibration of the production function using cost share data may require more than just the long-run average or median used in Section 2.4.1.

[31] As mentioned in Section 2.5.1, we could not conduct such thorough investigations using our best proxies due to their limited sample size, especially before the mid-1990s. Nonetheless, if anything, more recent data appear to corroborate these findings.

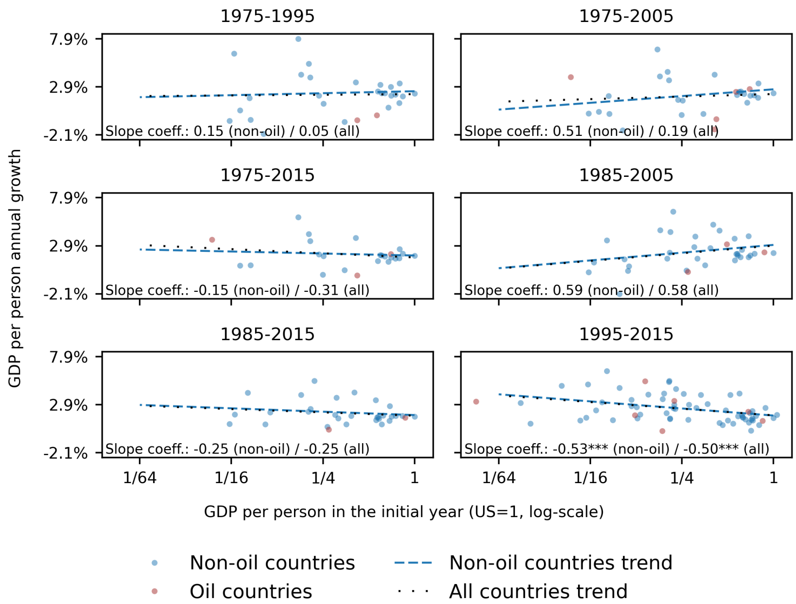


Figure 2.8 – Assessing convergence in income - data sets with greatest coverage, $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

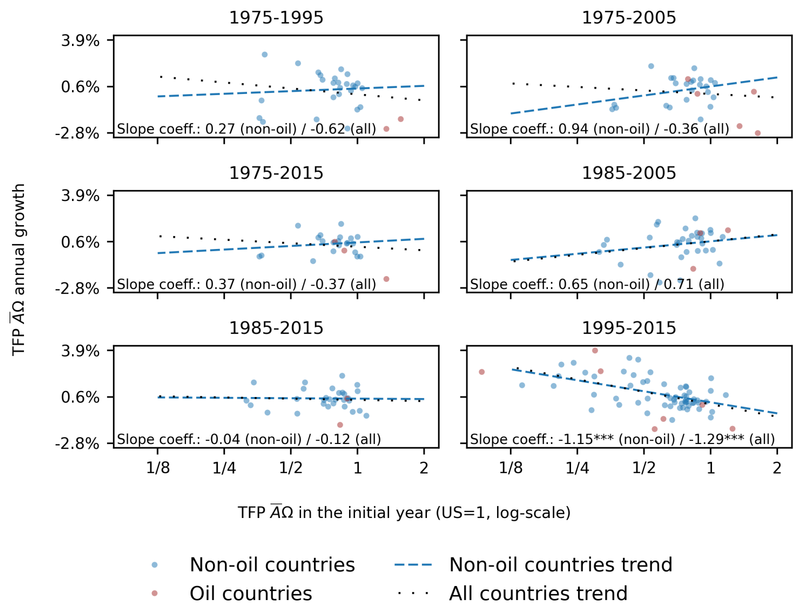


Figure 2.9 – Assessing convergence in TFP - data sets with greatest coverage, $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

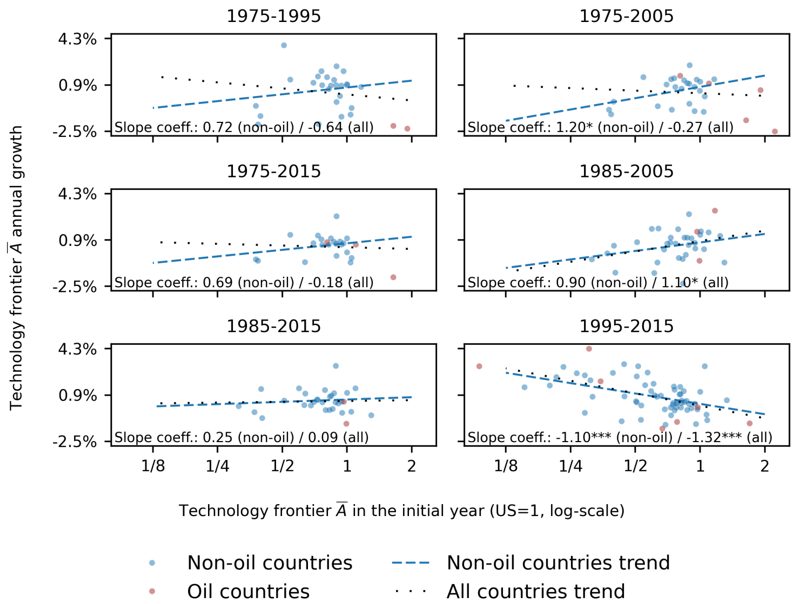


Figure 2.10 – Assessing convergence in technology frontier - data sets with greatest coverage, $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

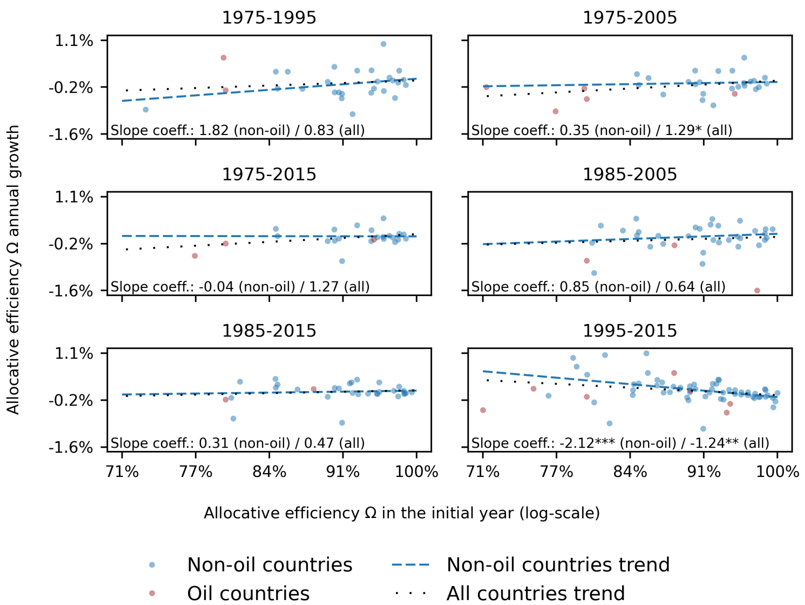


Figure 2.11 – Assessing convergence in allocative efficiency - data sets with greatest coverage, $k = -1$. *Note*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

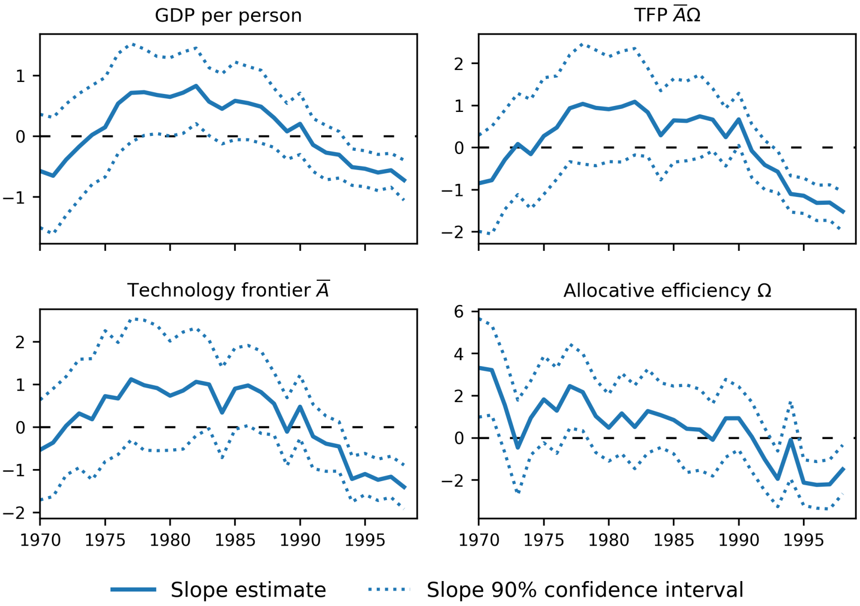


Figure 2.12 – Slope coefficient from regressing variable's annual growth over the next 20 years (in percent) on its level (in natural logarithm) - data sets with greatest coverage, $k = -1$.

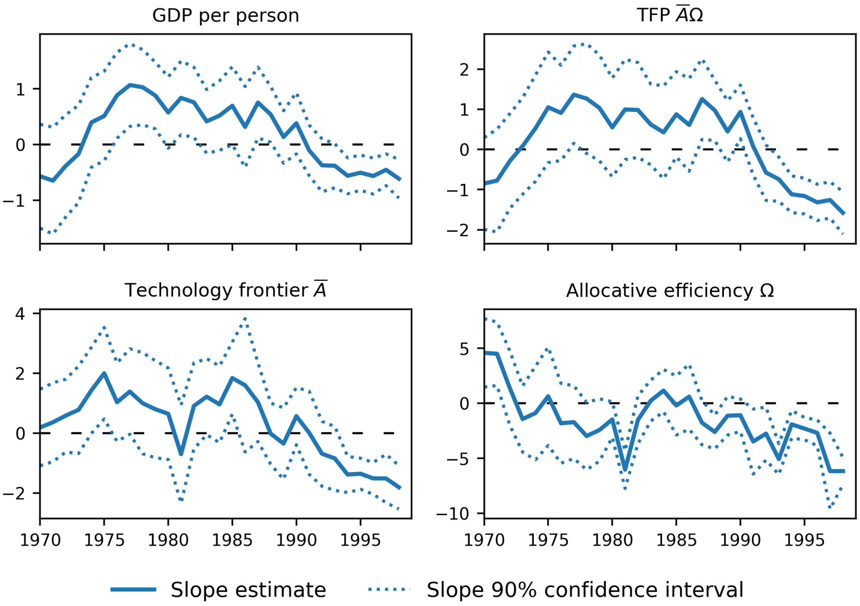


Figure 2.13 – Slope coefficient from regressing variable's annual growth over the next 20 years (in percent) on its level (in natural logarithm) - data sets with greatest coverage, $k = 3$.

Second, we add firm-specific wedges as a new source of firm heterogeneity, which could be helpful, for instance, to study the impact of size-dependent policies on misallocation. More precisely, we consider firm-specific tax rate over revenue $\tau_i = \tau(A_i)$. In this case, model equations change, and consequently, the conditions for the existence of a solution for the calibration algorithm of Section 2.3.3 are no longer valid. We do not obtain new such conditions but briefly comment on the empirical implementation of this model when (i) the function τ is known and (ii) the function τ is known except from a parameter, but the sales-weighted average tax rate $\tilde{\tau} \equiv \sum_{i=1}^{N} s_i \tau_i$ is observed.

2.7 Conclusion

Using a Cournot model, this study decomposes TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries from the Penn World Table 10.01. This decomposition enables a reexamination of key facts of economic growth. Our evaluation of the world income frontier, proxied by the US, reveals that changes in misallocation can significantly impact short-run growth. For example, during 2000-2007, the US witnessed notable technological improvement coupled with declining allocative efficiency, suggesting that the dot-com boom and advancements in IT led to productivity gains but concentrated in certain firms. On a more general note, the technology component seems to grow more steadily than the TFP itself, around 1% per year. Notable exceptions are the periods of 1954-1973 and 2000-2007 when technology contributed approximately 2% annually.

Turning to a global perspective, our analysis suggests that misallocation plays a significant role in explaining cross-country income differences. Including misallocation increases our first measure of success by raising the variance of the explained portion. We also observe an increased covariance between the explained portion and the income per unit of labor when allocative efficiency is considered. This explains why our second measure of success increases as well, even though the improvements are smaller in this case. Despite its significance, a considerable unexplained portion persists, constituting the majority of observed variability in most cases. Finally, we obtain limited support for the convergence hypothesis in income and either TFP component. Interestingly, the lack of convergence in allocative efficiency suggests that market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

Appendix of Chapter 2

2.A Derivation of the discrete model

We refrain from delving into households' behavior, as it is essentially irrelevant to our results; all we need is to assume that more consumption is always preferred to less. Consequently, the model focuses solely on the firms' side, where misallocation originates. Additionally, since firms' decisions are static in our model, we suppress the time subscript for notational simplicity.

2.A.1 Environment and technology

In a closed economy, N potential entrant firms produce a single good. The price elasticity of demand for this good is strictly negative, with its absolute value denoted by η, where $1 < \eta < \infty$. Since firms' goods are homogeneous, the aggregate output Y is

$$Y \equiv \sum_{i=1}^{N} Y_i \tag{2.1}$$

being Y_i the production of firm i, which is given by the Cobb-Douglas function

$$Y_i = A_i K_i^{\alpha} H_i^{1-\alpha} \tag{2.2}$$

where $K_i \geq 0$ is the stock of physical capital, $H_i \geq 0$ is the stock of human capital, and $A_i > 0$ is a productivity parameter, all for firm i, while $\alpha \in (0,1)$. In the following, let $\underline{\underline{A}} \equiv \min_i\{A_i\}$ and $\overline{A} \equiv \max_i\{A_i\}$ be the technology frontier of this economy, with $0 < \underline{\underline{A}} < \overline{A} < +\infty$.

2.A.2 Market competition and optimal decision

Firms engage in Cournot competition, meaning each firm chooses its output taking as given the output chosen by the other firms in the economy, as well as the wage $w > 0$ and the rental cost of physical capital $r > 0$. Formally, each firm $i \in \{1, 2, ..., N\}$ solves the profit maximization problem

$$\begin{aligned} \max_{Y_i} \quad & pY_i - wH_i - rK_i = (p - MC_i)\, Y_i \\ \text{s.t.} \quad & w > 0, r > 0, p = p(Y), Y_j \geq 0 \ \forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{aligned} \tag{2.3}$$

where p is the price of the good and $MC_i = \left(\frac{r}{\alpha}\right)^{\alpha} \left(\frac{w}{1-\alpha}\right)^{1-\alpha} \frac{1}{A_i}$ is the Cobb-Douglas marginal cost of firm i. The price is given by the inverse demand function $p(Y)$, with $-\left(\frac{\partial p}{\partial Y}\frac{Y}{p}\right)^{-1} \equiv \eta$.

The First-Order Condition (FOC) of this optimization problem is

$$\begin{aligned} 0 =& p\left(1 + \frac{\partial p}{\partial Y_i}\frac{Y_i}{p}\right) - MC_i = p\left(1 - \frac{1}{\eta_i}\right) - MC_i \\ p =& MC_i \frac{\eta_i}{\eta_i - 1} \end{aligned} \tag{2.4}$$

where $\eta_i \equiv -\left(\frac{\partial p}{\partial Y_i}\frac{Y_i}{p}\right)^{-1}$ is the absolute value of the price elasticity of demand faced by firm i. Being $s_i \equiv \frac{Y_i}{Y}$ the market share of firm i, note $\eta_i = \left(-\frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\frac{Y}{p}\frac{Y_i}{Y}\right)^{-1} = \frac{\eta}{s_i} > 1$. As a result, the markup of firm i is $\mu_i \equiv \frac{\eta_i}{\eta_i - 1} = \left(1 - \frac{s_i}{\eta}\right)^{-1}$.

The Second-Order Condition (SOC) is

$$\begin{aligned} 0 >& \frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\left(1 - \frac{s_i}{\eta}\right) - \frac{p}{\eta}\left(\frac{Y - Y_i\frac{\partial Y}{\partial Y_i}}{Y^2}\right) \\ 0 >& -\frac{p}{\eta Y}\left(1 - \frac{s_i}{\eta}\right) - \frac{p}{\eta Y}(1 - s_i) = -\frac{p}{\eta^2 Y}\left[\eta(2 - s_i) - s_i\right] \\ \eta >& \frac{s_i}{2 - s_i} \end{aligned} \tag{2.A.1}$$

which is satisfied for $\eta > 1$ as $\frac{s_i}{2-s_i}$ is strictly increasing in s_i and equals 1 for $s_i = 1$.

Therefore, Equation (2.4) represents firm i optimal decision as long as $\mu_i \geq 1 \leftrightarrow s_i \geq 0$, that is, for every active firm i.

2.A.3 Equilibrium allocation

Using (2.4) for any active firms i and j, note

$$MC_i(\eta - s_i)^{-1} = MC_j(\eta - s_j)^{-1}$$

$$s_i - s_j = \left(1 - \frac{MC_i}{MC_j}\right)(\eta - s_j) = \left(1 - \frac{A_j}{A_i}\right)(\eta - s_j) \tag{2.5}$$

As discussed in the main text, in the unique refined equilibrium, there exists a firm with productivity $\underline{A}$ serving as the cutoff for active firms, such that firm i is active if and only if $A_i \geq \underline{A}$. Given that, rewrite (2.5) as $\frac{s_i}{A_j} = \frac{\eta}{A_j} - \frac{1}{A_i}(\eta - s_j)$ and sum it in j over all active firms to obtain

$$\begin{aligned} s_i \mathrm{E}_a(1/A) =& \eta \mathrm{E}_a(1/A) - \frac{1}{N_a A_i}(N_a \eta - 1) \\ s(A_i) \equiv & s_i = \frac{1}{N_a \mathrm{E}_a(A_i/A)} + \eta\left(1 - \frac{1}{\mathrm{E}_a(A_i/A)}\right) \end{aligned} \tag{2.A.2}$$

where $\mathrm{E}_a(h(A)) \equiv \mathrm{E}(h(A)|A \geq \underline{A}) = \sum_{A \geq \underline{A}} h(A) \frac{g(A)}{1-G(\underline{A})}$ is the expected value of a function h over active firms under the empirical distribution, $g(A)$ is the empirical probability of A, $G(A) = \sum_{a<A} g(a)$ is the empirical cumulative distribution function, and $N_a \equiv N(1 - G(\underline{A}))$ is the number of active firms. As expected, these shares add to one over all active firms.

Given the free-entry assumption, we use $s(\underline{A}) \approx 0$, implying that, for $A_i \geq \underline{A}$,

$$s(A_i) \approx \eta\left(1 - \underline{A}/A_i\right) \tag{2.6}$$

$$\eta \approx \frac{1}{N_a\left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]} \tag{2.7}$$

where we obtain (2.6) by plugging $A_j = \underline{A}$ and $s_j \approx 0$ into (2.5), while we use $s(\underline{A}) \approx 0$ in (2.A.2) to get (2.7).

2.A.4 Aggregate productivity and misallocation

Given (2.1) and (2.2), note

$$Y = \sum_{i=1}^{N} A_i K_i^{\alpha} H_i^{1-\alpha} = \overline{A} \Omega K^{\alpha} H^{1-\alpha} \tag{2.8}$$

where $K \equiv \sum_{i=1}^{N} K_i$, $H \equiv \sum_{i=1}^{N} H_i$, and $\Omega \equiv \sum_{i=1}^{N} \theta_{Ki}^{\alpha} \theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right)$, with $\theta_{Ki} \equiv \frac{K_i}{K}$ and $\theta_{Hi} \equiv \frac{H_i}{H}$. Since (i) $\theta_{Ki}^{\alpha} \theta_{Hi}^{1-\alpha} \in [0, 1]$ and (ii) $\sum_{i=1}^{N} \theta_{Ki}^{\alpha} \theta_{Hi}^{1-\alpha} \leq 1$, we can conclude that $0 < \Omega \leq 1$.[32] Thus, to maximize Y given K and H, one should allocate all the inputs to the most productive firm, when $\Omega = 1$ and $Y = \overline{A}K^{\alpha}H^{1-\alpha}$, which is expected given our assumption of homogeneous goods. Hence, $\Omega \in (0, 1]$ gauges the distance from the optimal allocation of inputs, being our measure of allocative efficiency.

Firms use inputs optimally, taking the same inputs' rental prices as given. As a consequence, from the FOCs of active firm i cost minimization problem,

$$\frac{w}{MPH_i} = \frac{r}{MPK_i} \leftrightarrow \frac{wH_i}{(1-\alpha)Y_i} = \frac{rK_i}{\alpha Y_i} \leftrightarrow \frac{K_i}{H_i} = \frac{w}{r}\frac{\alpha}{1-\alpha} \tag{2.A.3}$$

where MPH_i and MPK_i are the marginal products of human and physical capital of active firm i, respectively. Thus, every active firm chooses the same physical-to-human capital ratio. Consequently, from (2.A.3), $K_i = \frac{K_j}{H_j}H_i$ for any firm i and active firm j, implying

$$\theta_{Kj} \equiv \frac{K_j}{K} = \frac{K_j}{\sum_{i=1}^{N} K_i} = \frac{K_j}{\sum_{i=1}^{N} \frac{K_j}{H_j} H_i} = \frac{H_j}{\sum_{i=1}^{N} H_i} = \frac{H_j}{L} \equiv \theta_{Hj} \tag{2.A.4}$$

which allows us to define $\theta_j \equiv \theta_{Hj} = \theta_{Kj}$ for every firm j. Given that, one can use (2.A.3) to get

$$\alpha = \frac{K_i r}{K_i r + H_i w} = \frac{\theta_i K r}{\theta_i K r + \theta_i H w} = \frac{Kr}{Kr + Hw} \tag{2.A.5}$$

meaning α equals the cost share of physical capital.

Using $\theta_{Ki} = \theta_{Hi} = \theta_i$, one can show that

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i K_i^{\alpha} H_i^{1-\alpha}}{\overline{A}\Omega K^{\alpha} H^{1-\alpha}} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{2.9}$$

$$\overline{A}\Omega = \frac{1}{\sum_{i=1}^{N} \frac{s(A_i)}{A_i}} \tag{2.10}$$

where we use (2.2) and (2.8) to get (2.9), while (2.10) is obtained by summing (2.9) over all firms.

[32] To see that $\sum_{i=1}^{N} \theta_{Ki}^{\alpha} \theta_{Hi}^{1-\alpha} \leq 1$, just use $\theta_{Ki}^{\alpha} \theta_{Hi}^{1-\alpha} = \exp\left(\alpha \ln(\theta_{Ki}) + (1-\alpha)\ln(\theta_{Hi})\right) \leq \exp\left[\ln(\alpha\theta_{Ki} + (1-\alpha)\theta_{Hi})\right] = \alpha\theta_{Ki} + (1-\alpha)\theta_{Hi}$, since ln is concave and exp is increasing.

Finally, plugging Equations (2.6) and (2.7) into (2.10),

$$\Omega \approx \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]} \tag{2.11}$$

which shows three interesting properties. First, it is easy to see that $\Omega \in (0,1]$, as it should be given (2.8). Second, $\Omega \to 1$ when $\underline{A} \to \overline{A}$, which is an expected result since with no productivity dispersion, any allocation of resources is optimal. After all, employing Equation (2.10),

$$1 = \sum_{i=1}^{N} s(A_i)(\overline{A}/\overline{A}) < \frac{1}{\Omega} = \sum_{i=1}^{N} s(A_i)(\overline{A}/A_i) < \sum_{i=1}^{N} s(A_i)(\overline{A}/\underline{A}) = \overline{A}/\underline{A} \tag{2.A.6}$$

implying $1 \leq \lim_{\underline{A}\to\overline{A}} \Omega^{-1} \leq \lim_{\underline{A}\to\overline{A}} \overline{A}/\underline{A} = 1$. Note this result relies solely on the assumption of homogeneous goods, not depending on the Cournot model. Third, the exit of less productive active firms leads to an improvement in Ω. To see that, we first assess the impact on allocative efficiency Ω resulting from the exit of only the least productive active firm from the market (say, because η increases). Let $\underline{A}+\delta$, with $\delta > 0$, be the second-lowest productivity among active firms in the initial equilibrium. In this proof, we use the subscript 0 to denote the initial equilibrium and 1 for the final one, implying $\mathrm{E}_{a_0}(h(A)) \equiv \mathrm{E}(h(A)|A \geq \underline{A})$ and $\mathrm{E}_{a_1}(h(A)) \equiv \mathrm{E}(h(A)|A \geq \underline{A}+\delta)$. Moreover, from Equation (2.11),

$$\begin{aligned}
\Omega_0 &\approx \frac{\mathrm{E}_{a_0}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}_{a_0}\left[(\underline{A}/A)(1-\underline{A}/A)\right]} = \frac{\mathrm{E}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)|A \geq \underline{A}\right]}{\mathrm{E}\left[(\underline{A}/A)(1-\underline{A}/A)|A \geq \underline{A}\right]} \\
\Omega_0 &\approx \frac{\mathrm{E}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)|A \geq \underline{A}+\delta\right]}{\mathrm{E}\left[(\underline{A}/A)(1-\underline{A}/A)|A \geq \underline{A}+\delta\right]} = \frac{\mathrm{E}_{a_1}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}_{a_1}\left[(\underline{A}/A)(1-\underline{A}/A)\right]}
\end{aligned} \tag{2.A.7}$$

$$\Omega_1 \approx \frac{\mathrm{E}_{a_1}\left[\left((\underline{A}+\delta)/\overline{A}\right)(1-(\underline{A}+\delta)/A)\right]}{\mathrm{E}_{a_1}\left[((\underline{A}+\delta)/A)(1-(\underline{A}+\delta)/A)\right]} \tag{2.A.8}$$

where we use $1-\underline{A}/\underline{A} = 0$ in the second line. Therefore, $\Delta\Omega \equiv \Omega_1 - \Omega_0$ is

$$\begin{aligned}
\Delta\Omega &\approx \frac{\mathrm{E}_{a_1}\left[\left((\underline{A}+\delta)/\overline{A}\right)(1-(\underline{A}+\delta)/A)\right]}{\mathrm{E}_{a_1}\left[((\underline{A}+\delta)/A)(1-(\underline{A}+\delta)/A)\right]} - \frac{\mathrm{E}_{a_1}\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}_{a_1}\left[(\underline{A}/A)(1-\underline{A}/A)\right]} \\
\Delta\Omega &\approx \frac{\mathrm{E}_{a_1}\left[(\underline{A}+\delta)\left(\frac{A-(\underline{A}+\delta)}{\overline{A}A}\right)\right]}{\mathrm{E}_{a_1}\left[(\underline{A}+\delta)\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\right]} - \frac{\mathrm{E}_{a_1}\left[\underline{A}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\right]}{\mathrm{E}_{a_1}\left[\underline{A}\left(\frac{A-\underline{A}}{A^2}\right)\right]} = \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A} - \frac{\delta}{\overline{A}A}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2} - \frac{\delta}{A^2}\right)} - \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)}
\end{aligned}$$

$$\Delta\Omega \approx \frac{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A} - \frac{\delta}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right) - \mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2} - \frac{\delta}{A^2}\right)}{\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2} - \frac{\delta}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)}$$

$$\Delta\Omega \approx \frac{-\mathrm{E}_{a_1}\left(\frac{\delta}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right) + \mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{\overline{A}A}\right)\mathrm{E}_{a_1}\left(\frac{\delta}{A^2}\right)}{\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)}$$

$$\Delta\Omega \approx \frac{\mathrm{E}_{a_1}\left(1-\frac{\underline{A}}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right) - \mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A} - \frac{\underline{A}}{A^2}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)}$$

$$\Delta\Omega \approx \frac{-\mathrm{E}_{a_1}\left(\frac{1}{A}\right)^2 + \underline{A}\mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right) + \mathrm{E}_{a_1}\left(\frac{1}{A^2}\right) - \underline{A}\mathrm{E}_{a_1}\left(\frac{1}{A}\right)\mathrm{E}_{a_1}\left(\frac{1}{A^2}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)}$$

$$\Delta\Omega \approx \frac{\mathrm{Var}_{a_1}\left(\frac{1}{A}\right)}{(\overline{A}/\delta)\mathrm{E}_{a_1}\left(\frac{A-(\underline{A}+\delta)}{A^2}\right)\mathrm{E}_{a_1}\left(\frac{A-\underline{A}}{A^2}\right)} \tag{2.A.9}$$

where $\mathrm{Var}_{a_1}(h(A)) \equiv \mathrm{Var}(h(A)|A \geq \underline{A}+\delta) = \mathrm{E}_{a_1}(h(A)^2) - [\mathrm{E}_{a_1}(h(A))]^2$ is the variance of $h(A)$ over active firms under the empirical distribution in the final equilibrium, for any function h. Note that our analysis implicitly assumes the presence of productivity dispersion in the initial equilibrium. Within the initial set of active firms, the second-lowest productivity level is $\underline{A}+\delta$, which is strictly greater than the lowest level ($\underline{A}$) as $\delta > 0$. As a result, productivity dispersion should also be present in the final equilibrium. After all, if this were not the case, $\mathrm{E}_{a_1}(\underline{A}/A) = 1$, implying from (2.7) that either $N_{a_1} = +\infty$ or $\eta_1 = +\infty$. While $\eta_1 = +\infty$ is ruled out by assumption, the scenario of $N_{a_1} = +\infty$ cannot hold since there is productivity dispersion in the initial equilibrium. All in all, we can conclude that $\mathrm{Var}_{a_1}(1/A) > 0$ and consequently $\Delta\Omega > 0$. Applying this result iteratively, we would conclude that the exit of less productive active firms improves the allocative efficiency Ω.

2.A.5 Average markup

Using this model, we can also compute the cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N}\left(\frac{H_i w + K_i r}{Hw+Kr}\right)\mu_i = \sum_{i=1}^{N}\theta_i\mu_i$ through

$$\mu = \sum_{i=1}^{N}\left[\overline{A}\Omega\frac{s(A_i)}{A_i}\right]\left(\frac{p}{MC_i}\right) \approx \sum_{i=1}^{N}\left[\overline{A}\Omega\frac{s(A_i)}{A_i}\right]\left[\frac{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{\underline{A}}}{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{A_i}}\right] = \frac{\overline{A}\Omega}{\underline{A}} \tag{2.12}$$

where we use (2.9), (2.4), that the least productive active firm has markup approximately equal to one, and the Cobb-Douglas marginal cost function.

2.B Derivation of the continuous model

Given the similarity to the discrete model, we refrain from presenting the derivation in detail. Instead, our focus is on emphasizing the main differences from the discrete case. To derive the model, the SOC of firms' profit maximization must hold. In the discrete case, one can easily see that this condition is satisfied for $\eta > 1$. However, in the continuous model, such evaluation is less straightforward, as it requires considering the equilibrium values of s_i. As a consequence, in deriving the model, we simply assume this condition is met. We subsequently validate this claim in Appendix 2.B.6 using the model solution for s_i, demonstrating that the SOC holds if $\eta > 0$ or if $q \in (0, 1]$ is low.

2.B.1 Environment and technology

The absolute value of the price elasticity of demand is η, with $0 < \eta < +\infty$. Since now there is a continuum of firms $i \in [0, N]$, the aggregate output is given by an integral instead of a sum:

$$Y \equiv \int_0^N Y_i di \tag{2.1c}$$

where Y_i is still given by (2.2).

2.B.2 Market competition and optimal decision

Firms' problem continues to be (2.3). Consequently, the FOC is still given by (2.4), but now $\eta_i = \frac{\eta/q}{s_i}$ since $\partial Y/\partial Y_i = q \in (0, 1]$ instead of $\partial Y/\partial Y_i = 1$. This means the markup of firm i becomes $\mu_i = \left(1 - \frac{s_i}{\eta/q}\right)^{-1} = \frac{\eta/q}{\eta/q - s_i}$. Similarly, the SOC is also different, given by

$$0 > \frac{\partial p}{\partial Y}\frac{\partial Y}{\partial Y_i}\left(1 - \frac{s_i}{\eta/q}\right) - \frac{p}{\eta/q}\left(\frac{Y - Y_i\frac{\partial Y}{\partial Y_i}}{Y^2}\right)$$

$$0 > -\frac{p}{(\eta/q)Y}\left(1-\frac{s_i}{\eta/q}\right)-\frac{p}{(\eta/q)Y}(1-s_iq) = -\frac{p}{(\eta^2/q)Y}\left[2\eta - s_iq(1+\eta)\right]$$

$$s_iq < \frac{2\eta}{1+\eta} \tag{2.A.1c}$$

For now, simply assume the SOC holds, implying (2.4) represents the optimal decision for all active firms, that is, for every firm i such as $s_i \geq 0$ or, equivalently, $\mu_i \geq 1$.

2.B.3 Equilibrium allocation

Equation (2.5) is no longer valid. However, from (2.4) with $\eta_i = \frac{\eta/q}{s_i}$, one can easily show that

$$s_i - s_j = \left(1-\frac{A_j}{A_i}\right)(\eta/q - s_j) \tag{2.5c}$$

which defines s_i as a strictly increasing function of A_i, since $\eta/q - s_j > 0$ as $\mu_j = \frac{\eta/q}{\eta/q - s_j} > 1$ for any active firm j. As before, if some firms may be inactive, we seek the unique equilibrium in which a firm i is active if and only if $A_i \geq \underline{A}$ for some productivity cutoff $\underline{A}$.

Given that, one can use (2.5c) to obtain, analogously to (2.A.2),

$$s(A_i) = \frac{1}{N_a \mathrm{E}_a(A_i/A)} + (\eta/q)\left(1-\frac{1}{\mathrm{E}_a(A_i/A)}\right) \tag{2.A.2c}$$

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|A \geq \underline{A}\right) = \int_{\underline{A}}^{\overline{A}} h(A)\frac{g(A)}{1-G(\underline{A})}dA$ is the expected value of a function h over active firms under the empirical distribution, $g(A) > 0$ is the empirical density of A, and $G(A) = \int_{\underline{\underline{A}}}^{A} g(a)da$ is the empirical cumulative distribution function. As before, $N_a \equiv N\left(1 - G(\underline{A})\right)$ is the number of active firms. Using (2.A.2c), let us show that not all firms can be active simultaneously if N is sufficiently large. Assume, by contradiction, that all firms are active, with $N_a = N \to +\infty$. From (2.A.2c), it is easy to see that $s(\underline{A}) < 0$ under such circumstance unless $E_a(\underline{A}/A) \to 1$ and thus $\underline{A} \to \overline{A}$, when you would get the perfect competition case. However, this could not hold as all firms are active and $A_i \in [\underline{\underline{A}}, \overline{A}]$, with $\underline{\underline{A}} < \overline{A}$.

Therefore, assuming that N is sufficiently large, some low-productivity firms would be inactive. Consequently, $s(\underline{A}) = 0$. To see this last result, it is sufficient to show that such $\underline{A}$ exists. After all, on the one hand, if it exists, $s(\underline{A}) > 0$ cannot be an equilibrium given our assumption of free entry. On the other hand, $s(\underline{A}) < 0$ is never an equilibrium due to free exit. A productivity level $\underline{A}$ such as $s(\underline{A}) = 0$ exists because (i) $s(\underline{A}) < 0$ when $\underline{A} \to \underline{\underline{A}}$, (ii) $s(\underline{A}) > 0$ when $\underline{A} \to \overline{A}$, and (iii) $s(\underline{A})$ is continuous in $\underline{A}$. The first result holds because s_i is strictly increasing in A_i and, by assumption, the mass N of firms is large, meaning that not all firms can be active simultaneously. The last two results can be obtained from (2.A.2c). From this expression, $s(\underline{A}) \to 1/N_a = 1/[N(1 - G(\underline{A}))] \to +\infty$ when $\underline{A} \to \overline{A}$, yielding the second result. The third result holds as both $N_a \equiv N(1 - G(\underline{A}))$ and $\mathrm{E}_a(1/A)$ are continuous functions of $\underline{A}$, since $\mathrm{E}_a(1/A)$ is differentiable in $\underline{A}$ (Proposition 2.D.1).

Finally, plugging $s(\underline{A}) = 0$ respectively into (2.5c) and (2.A.2c), note

$$s(A_i) = (\eta/q)\,(1 - \underline{A}/A_i) \tag{2.6c}$$

$$\eta/q = \frac{1}{N_a\,[1 - \mathrm{E}_a\,(\underline{A}/A)]} \tag{2.7c}$$

which are the continuous versions of (2.6) and (2.7), respectively.

2.B.4 Aggregate productivity and misallocation

Equation (2.8) remains valid with $K \equiv \int_0^N K_i di$, $H \equiv \int_0^N H_i di$, and $\Omega \equiv \int_0^N \theta_{Ki}^{\alpha}\theta_{Hi}^{1-\alpha}\left(A_i/\overline{A}\right) di$. Equations (2.A.3), (2.A.4), (2.A.5), and (2.9) also remain valid. Naturally, (2.10) should be replaced by

$$\overline{A}\Omega = \frac{1}{\int_0^N \frac{s(A_i)}{A_i} di} \tag{2.10c}$$

Finally, allocative efficiency Ω continues to be given by (2.11), but now holding *exactly*. Thus, it shows similar properties. First, as can be easily seen, $\Omega \in (0, 1]$. Second, $\Omega \to 1$ when $\underline{A} \to \overline{A}$, since, from (2.10c),

$$1 = \int_0^N s(A_i)(\overline{A}/\overline{A})di < \frac{1}{\Omega} = \int_0^N s(A_i)(\overline{A}/A_i)di < \int_0^N s(A_i)(\overline{A}/\underline{A})di = \overline{A}/\underline{A} \tag{2.A.6c}$$

Third, Ω is strictly increasing in $\underline{A}$, since, from Equation (2.11) holding exactly,

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\left[-\mathrm{E}_a(1/A)-\underline{A}\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}\right]\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}-\frac{\left[1-\underline{A}\mathrm{E}_a(1/A)\right]\left[\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}-E(1/A^2)-\underline{A}\frac{\partial\mathrm{E}_a(1/A^2)}{\partial\underline{A}}\right]}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}\left[\underline{A}^2\mathrm{E}_a(1/A^2)-1\right]+\underline{A}\frac{\partial\mathrm{E}_a(1/A^2)}{\partial\underline{A}}\left[1-\underline{A}\mathrm{E}_a(1/A)\right]}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}-\frac{\mathrm{E}_a(1/A)\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]-\left[1-\underline{A}\mathrm{E}_a(1/A)\right]E(1/A^2)}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right]-1\right\}-\underline{A}\frac{\partial\mathrm{E}_a(1/A^2)}{\partial\underline{A}}\left[\mathrm{E}_a(\underline{A}/A)-1\right]}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}+\frac{\mathrm{E}_a(1/A^2)-\left[\mathrm{E}_a(1/A)\right]^2}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\tilde{g}(\underline{A})\left[\mathrm{E}_a\left(1/A\right)-(1/\underline{A})\right]\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right]-1\right\}}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}+\frac{-\underline{A}\tilde{g}(\underline{A})\left[\mathrm{E}_a\left(1/A^2\right)-(1/\underline{A}^2)\right]\left[\mathrm{E}_a(\underline{A}/A)-1\right]+\mathrm{Var}_a(1/A)}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\left[\frac{\tilde{g}(\underline{A})}{\underline{A}}\right]\left[\mathrm{E}_a\left(\underline{A}/A\right)-1\right]\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right]-1\right\}}{\overline{A}\left[\mathrm{E}_a(1/A)-\underline{A}\mathrm{E}_a(1/A^2)\right]^2}+\frac{-\left[\frac{\tilde{g}(\underline{A})}{\underline{A}}\right]\left\{\mathrm{E}_a\left[(\underline{A}/A)^2\right]-1\right\}\left[\mathrm{E}_a(\underline{A}/A)-1\right]+\mathrm{Var}_a(\underline{A}/A)}{\overline{A}\left\{\mathrm{E}_a(\underline{A}/A)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial\Omega}{\partial\underline{A}}=\frac{\mathrm{Var}_a(\underline{A}/A)}{\overline{A}\left\{\mathrm{E}_a(\underline{A}/A)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]\right\}^2}$$

where $\mathrm{Var}_a(h(A))\equiv\mathrm{E}_a\left(h(A)^2\right)-\left[\mathrm{E}_a(h(A))\right]^2$ and we use Proposition 2.D.1 to get $\frac{\partial\mathrm{E}_a(1/A)}{\partial\underline{A}}$ and $\frac{\partial\mathrm{E}_a(1/A^2)}{\partial\underline{A}}$ in the fourth line. Hence, $\frac{\partial\Omega}{\partial\underline{A}}>0$, since $\mathrm{Var}_a\left(1/A\right)>0$ given (2.7c) with $0<\eta<+\infty$ and $q\in(0,1]$.

2.B.5 Average markup

Equation (2.12) is now *exactly* valid.

2.B.6 Assessing the SOC for profit maximization

As our final task, we assess if the SOC for profit maximization (2.A.1c) holds. Note $\eta > 1$ is not sufficient because, with a continuum of firms, s_i is a density function, and consequently, it may be strictly greater than one. Since s_i is an endogenous variable, let us evaluate this condition in the model, using $s_i q = \eta\,(1 - \underline{A}/A_i)$ from Equation (2.6c).

Formally, we need to demonstrate that $\frac{2\eta}{1+\eta} > s_i q = \eta\,(1 - \underline{A}/A_i)$ for every active firm i. We address this issue under two cases: (i) $\eta \in (0, 1]$ and (ii) $\eta > 1$. On the one hand, since $\eta \in (0, 1] \to \frac{2\eta}{1+\eta} \geq \eta$ and $\eta > s_i q$ as $\mu_i = \frac{\eta}{\eta - s_i q} > 1$ for any active firm i, the SOC holds for any $\eta, q \in (0, 1]$.[33] On the other hand, the SOC holds for $\eta \approx 1$, $\eta > 1$, since it holds for $\eta = 1$ and we are just dealing with continuous functions of η, but it is not fulfilled for high enough η. To get this last result, use (2.7c) to see that, when $\eta \to +\infty$, $\underline{A} \to \overline{A}$ and thus $\eta \mathrm{E}_a(1 - \underline{A}/A) = \frac{q}{N_a} = \frac{q}{N(1-G(\underline{A}))} \to +\infty$. As a consequence, given that $(1 - \underline{A}/\overline{A}) \geq \mathrm{E}_a(1 - \underline{A}/A)$, $\eta(1 - \underline{A}/\overline{A}) \to +\infty$ when $\eta \to +\infty$. This implies the SOC will not hold for the most productive firms under high η since the LHS of (2.A.1c) is always lower than 2. Moreover, with $\eta > 1$, the SOC holds for low $q > 0$, since $s_i q = \eta\,(1 - \underline{A}/A_i) \to 0$ when $q \to 0^+$ as, from (2.7c), $\underline{A} \to \overline{A}$ under such condition.

In short, the SOC holds for low $\eta > 0$ or low $q \in (0, 1]$. Note that by choosing both η and q, we can simultaneously fulfill such condition and choose the adjusted elasticity of demand η/q (e.g., set $\eta = 1$ and choose q to get the desired η/q). However, in empirical applications, calibrating η/q is not necessarily needed as several key variables, such as allocative efficiency Ω, can be computed without knowing it. In this context, one may only assume $\eta > 0$ and $q \in (0, 1]$ are such that $\frac{2\eta}{1+\eta} > s_i q$ for every active firm i. We adopt this approach in the main text.

[33] Through analogous reasoning, it can be shown that $\eta \in (0, 1]$ is also suitable for the discrete model. Therefore, while $\eta > 1$ is a sufficient condition there, it is not necessary.

2.C Microfoundation of the price elasticity of demand

To microfoundate the price elasticity of demand η, rather than relying on a single-sector (or single-good) economy, we examine a multiple-sector economy but that can be represented within a single-sector environment. Under this new setup, our model would apply to that representative sector. However, in this case, the price elasticity of demand is not an ad hoc parameter; instead, it is equal to the elasticity of substitution across sectors.

Formally, suppose a representative perfectly competitive firm produces a homogeneous final good Y using inputs Y_s from a continuum of sectors through the CES production function

$$Y = \left(\int_0^1 Y_s^{\frac{\sigma-1}{\sigma}} ds \right)^{\frac{\sigma}{\sigma-1}} \tag{2.C.1}$$

where σ is the elasticity of substitution across sectors $s \in [0,1]$. Therefore, if the final good is the numeraire, this representative firm solves the profit maximization problem

$$\begin{aligned} \max_{\{Y_s\}_{s\in[0,1]}} \quad & \left(\int_0^1 Y_s^{\frac{\sigma-1}{\sigma}} ds \right)^{\frac{\sigma}{\sigma-1}} - \int_0^1 Y_s p_s ds \\ \text{s.t.} \quad & p_s > 0, \quad \forall s \in [0,1] \end{aligned} \tag{2.C.2}$$

with p_s representing the price of goods within sector s, which are homogeneous across all firms in that sector. The FOCs of this problem are

$$p_s = Y^{1/\sigma} Y_s^{-1/\sigma}, \quad \forall s \in [0,1] \tag{2.C.3}$$

In each sector, intermediate firms make their decisions taking the price and quantity of the final good as given.[34] As a result, being η_s the absolute value of the price elasticity of demand for the good of sector s, using (2.C.3), we can see that

$$\eta_s \equiv - \left(\frac{\partial p_s}{\partial Y_s} \frac{Y_s}{p_s} \right)^{-1} = \sigma \left(Y^{1/\sigma} Y_s^{-1/\sigma - 1} \frac{Y_s}{Y^{1/\sigma} Y_s^{-1/\sigma}} \right)^{-1} = \sigma \tag{2.C.4}$$

[34] This assumption is also pivotal if one aims to specify and solve a full macroeconomic model because it overcomes the technical problems associated with embedding oligopoly models into general equilibrium frameworks (Neary, 2010). This represents another advantage of this alternative interpretation of the model.

Hence, the price elasticity of demand is the same for all sectors, given by the elasticity of substitution σ. If we further assume that firms' technology, the number of firms, and the distribution of productivity are identical across sectors, each sector would face the same problem, and consequently $Y_s = Y_{\tilde{s}}$ for any sectors $s, \tilde{s} \in [0,1]$. Plugging it into (2.C.1),

$$Y = \left(\int_0^1 Y_{\tilde{s}}^{\frac{\sigma-1}{\sigma}} ds\right)^{\frac{\sigma}{\sigma-1}} = Y_{\tilde{s}} \tag{2.C.5}$$

implying that we can treat the economy as if there were a single sector, which we analyze through the lens of a static Cournot model.

2.D Conditions for the calibration algorithm to work properly

Owing to the simplicity of our model, we can establish necessary and sufficient conditions for the calibration algorithm of Section 2.3.3 to work properly, achieving an exact match of both target moments. Basically, we need to show that a unique solution exists for its first-step problem or, equivalently, that (2.14) implicitly defines $\tilde{A} \equiv \overline{A}/\underline{A}$ as a well-defined function of μ. Initially, we derive some general results by examining an arbitrary continuous truncated distribution. The only requirement is that its density can be expressed as a truncation of another one from above. These general findings provide a framework for establishing conditions applicable to any such distribution. In the latter part, we utilize this framework to delineate the specific conditions pertaining to the Pareto case. Additionally, in the final proposition of this section, we assess what happens with the estimated Ω under the Pareto distribution when $k \to -\infty$.

In the following, consider μ as given in (2.14), that is,

$$\mu = \frac{1 - \mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left(\underline{A}/A\right) - \mathrm{E}_a\left[(\underline{A}/A)^2\right]} = \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left(1 - \underline{A}/A\right) - \mathrm{E}_a\left[(1 - \underline{A}/A)^2\right]} \tag{2.14}$$

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A) | \underline{A} \leq A \leq \overline{A}\right)$ for any function h.

2.D.1 Arbitrary distribution

Assume $A \in [\underline{A}, \overline{A}]$ is a continuous variable, $0 < \underline{A} < \overline{A} < +\infty$, whose density and cumulative distribution function are g and G, respectively. Let $\tilde{g}(A) \equiv$

$\frac{g(A)}{1-G(\underline{A})} > 0$, $\underline{A} \in (\underline{\underline{A}}, \overline{A})$, the density function of $A \in [\underline{A}, \overline{A}]$, with cumulative distribution $\tilde{G}$. Let $\hat{g}$ be another density of A, but defined over the support $A \in [\underline{A}, A_h]$, $A_h > \overline{A}$, possibly with $A_h \to +\infty$. This density does not depend on $\overline{A}$ and has cumulative distribution function $\hat{G}$. Moreover, it satisfies $\tilde{g}(A) = \frac{\hat{g}(A)}{\hat{G}(\overline{A})}$, meaning $\tilde{g}$ is a truncation of $\hat{g}$ from above.

Proposition 2.D.1 *Let h be a function of A for which $\mathrm{E}_a(h(A))$ is well defined. In this case,*

1. *If* $\frac{\partial h(A)}{\partial \underline{A}} = 0$, $\frac{\partial \mathrm{E}_a(h(A))}{\partial \underline{A}} = \tilde{g}(\underline{A})\left[\mathrm{E}_a(h(A)) - h(\underline{A})\right]$.
2. *If* $\frac{\partial h(A)}{\partial \overline{A}} = 0$, $\frac{\partial \mathrm{E}_a(h(A))}{\partial \overline{A}} = \tilde{g}(\overline{A})\left[h(\overline{A}) - \mathrm{E}_a(h(A))\right]$.

Proof. Let h be a function of A for which $\mathrm{E}_a(h(A))$ is well defined. If $\frac{\partial h(A)}{\partial \underline{A}} = 0$,

$$\frac{\partial \mathrm{E}_a(h(A))}{\partial \underline{A}} = \frac{\partial \left[\frac{\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA}{1-G(\underline{A})}\right]}{\partial \underline{A}} = \frac{\left[\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA\right] g(\underline{A}) - h(\underline{A})g(\underline{A})\left(1-G(\underline{A})\right)}{\left[1-G(\underline{A})\right]^2}$$
$$\frac{\partial \mathrm{E}_a(h(A))}{\partial \underline{A}} = \tilde{g}(\underline{A})\left[\mathrm{E}_a(h(A)) - h(\underline{A})\right]$$

where we use $\frac{\partial h(A)}{\partial \underline{A}} = 0$ to get $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)g(A)dA\right]}{\partial \underline{A}} = -h(\underline{A})g(\underline{A})$. If $\frac{\partial h(A)}{\partial \overline{A}} = 0$,

$$\frac{\partial \mathrm{E}_a(h(A))}{\partial \overline{A}} = \frac{\partial \left[\frac{\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA}{\hat{G}(\overline{A})}\right]}{\partial \overline{A}} = \frac{h(\overline{A})\hat{g}(\overline{A})\hat{G}(\overline{A}) - \left[\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA\right]\hat{g}(\overline{A})}{\hat{G}(\overline{A})^2}$$
$$\frac{\partial \mathrm{E}_a(h(A))}{\partial \overline{A}} = \tilde{g}(\overline{A})\left[h(\overline{A}) - \mathrm{E}_a(h(A))\right]$$

where we use $\frac{\partial h(A)}{\partial \overline{A}} = 0$ to get $\frac{\partial\left[\int_{\underline{A}}^{\overline{A}} h(A)\hat{g}(A)dA\right]}{\partial \overline{A}} = h(\overline{A})\hat{g}(\overline{A})$. ∎

Proposition 2.D.2 $\frac{\partial \mu}{\partial \overline{A}} > 0$.

Proof. From Equation (2.14),

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}} \left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2} - \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\left\{\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}} - \frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}}\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}} - \frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}}\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)\tilde{g}(\overline{A})\left\{(1-\underline{A}/\overline{A})^2 - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2} - \frac{\tilde{g}(\overline{A})\left[(1-\underline{A}/\overline{A}) - \mathrm{E}_a\left(1-\underline{A}/A\right)\right]\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\tilde{g}(\overline{A})(1-\underline{A}/\overline{A})^2\mathrm{E}_a\left(1-\underline{A}/A\right) - \tilde{g}(\overline{A})(1-\underline{A}/\overline{A})\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

$$\frac{\partial \mu}{\partial \overline{A}} = \frac{\tilde{g}(\overline{A})(1-\underline{A}/\overline{A})\mathrm{E}_a\left[\left(1-\underline{A}/A\right)\left(\underline{A}/A - \underline{A}/\overline{A}\right)\right]}{\left\{\mathrm{E}_a\left(1-\underline{A}/A\right) - \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]\right\}^2}$$

where we use Proposition 2.D.1 to get $\frac{\partial \mathrm{E}_a(1-\underline{A}/A)}{\partial \overline{A}}$ and $\frac{\partial \mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{\partial \overline{A}}$ in the third line. Therefore, $\frac{\partial \mu}{\partial \overline{A}} > 0$ as $\left(1-\underline{A}/A\right)\left(\underline{A}/A - \underline{A}/\overline{A}\right) > 0$ for $A \in (\underline{A}, \overline{A})$. ■

Proposition 2.D.3 $\lim_{\overline{A} \to \underline{A}^+} \mu = 1$.

Proof. From Equation (2.14),

$$1 = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)}{\mathrm{E}_a\left[(1-\underline{A}/A)(\underline{A}/\underline{A})\right]} \leq \mu \leq \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)}{\mathrm{E}_a\left[(1-\underline{A}/A)(\underline{A}/\overline{A})\right]} = \overline{A}/\underline{A}$$

Therefore, $1 \leq \mu \leq \overline{A}/\underline{A}$, implying $\lim_{\overline{A} \to \underline{A}^+} \mu = 1$. ■

Proposition 2.D.4 *$\overline{A}$, $\overline{A} > \underline{A}$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$.*

Proof. First, μ is continuous and strictly increasing in $\overline{A}$, $\overline{A} > \underline{A}$, as it is differentiable with $\frac{\partial \mu}{\partial \overline{A}} > 0$ (Proposition 2.D.2), implying it is an one-to-one function. Second, since $\lim_{\overline{A}\to\underline{A}^+} \mu = 1$ (Proposition 2.D.3), the image of μ over $\overline{A} \in (\underline{A}, +\infty)$ is $\left(1, \lim_{\overline{A}\to+\infty} \mu\right)$ as μ is continuous and strictly increasing in $\overline{A}$. As a consequence of these two results, $\overline{A}$, $\overline{A} > \underline{A}$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$. ■

2.D.2 Pareto distribution of firm productivity

If $A \in \left[\underline{\underline{A}}, \overline{A}\right] \in (0, +\infty)$ is truncated Pareto distributed with shape parameter $k \neq 0$, its density is $g(A) = k \left(\frac{\underline{\underline{A}}^k \overline{A}^k}{\overline{A}^k - \underline{\underline{A}}^k}\right) A^{-k-1}$ and its cumulative distribution function is $G(A) = \frac{\overline{A}^k - \underline{\underline{A}}^k \overline{A}^k A^{-k}}{\overline{A}^k - \underline{\underline{A}}^k}$. Note this density is well defined for any $k \neq 0$ as $g(A) > 0 \; \forall A \in \left[\underline{\underline{A}}, \overline{A}\right]$ and $G(\overline{A}) = 1$. Moreover, it is easy to see $A \in \left[\underline{A}, \overline{A}\right]$ has also a truncated Pareto distribution with parameter $k \neq 0$, since its density is $\tilde{g}(A) \equiv \frac{g(A)}{1-G(\underline{A})} = k \left(\frac{\underline{A}^k \overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1}$. Consider in the following $\tilde{A} \equiv \overline{A}/\underline{A} > 1$.

Proposition 2.D.5 *For $k \neq 0$ and $j \in \mathbb{N} \setminus \{0\}$,*

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \begin{cases} \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right) & , \text{ if } k+j \neq 0 \\ \left(\frac{k\tilde{A}^k}{\tilde{A}^k-1}\right) \ln \tilde{A} & , \text{ if } k+j = 0 \end{cases}$$

Proof. Let $k \neq 0$ and $j \in \mathbb{N} \setminus \{0\}$. From the truncated Pareto density, if $k + j \neq 0$,

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \int_{\underline{A}}^{\overline{A}} k \left(\frac{\underline{A}^{k+j}\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-k-1-j} dA = k \left(\frac{\underline{A}^{k+j}\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right)\left(\frac{\overline{A}^{-k-j} - \underline{A}^{-k-j}}{-k-j}\right)$$

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = k \left(\frac{\tilde{A}^k}{\tilde{A}^k - 1}\right)\left(\frac{\tilde{A}^{-k-j} - 1}{-k-j}\right) = \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j} - 1}{\tilde{A}^{k+j} - \tilde{A}^j}\right)$$

while, if $k + j = 0$, $\mathrm{E}_a\left((\underline{A}/A)^j\right) = \int_{\underline{A}}^{\overline{A}} k \left(\frac{\overline{A}^k}{\overline{A}^k - \underline{A}^k}\right) A^{-1} dA = \left(\frac{k\tilde{A}^k}{\tilde{A}^k - 1}\right) \ln \tilde{A}$. ■

Proposition 2.D.6 $\mu = \begin{cases} \left(\frac{k+2}{k}\right) \frac{\tilde{A}^2\left(\tilde{A}^k-1\right)-k\tilde{A}\left(\tilde{A}-1\right)}{\tilde{A}\left(\tilde{A}^{k+1}-1\right)-(k+1)\left(\tilde{A}-1\right)} & , \textit{if } k \neq 0, -1, -2 \\ \frac{(\tilde{A}-1)-\ln\tilde{A}}{\ln\tilde{A}-\left(\frac{\tilde{A}-1}{\tilde{A}}\right)} & , \textit{if } k = -1 \\ \left(\frac{1}{2}\right) \frac{(\tilde{A}-1)^2}{(\tilde{A}-1)-\ln\tilde{A}} & , \textit{if } k = -2 \end{cases}$

Proof. If $k = -1$, using Proposition 2.D.5 in Equation (2.14),

$$\mu = \frac{1 - \frac{\ln\tilde{A}}{\tilde{A}-1}}{\frac{\ln\tilde{A}}{\tilde{A}-1} - \left(\frac{\tilde{A}-1}{\tilde{A}^2-\tilde{A}}\right)} = \frac{(\tilde{A}-1) - \ln\tilde{A}}{\ln\tilde{A} - \left(\frac{\tilde{A}-1}{\tilde{A}}\right)}$$

If $k = -2$, from Proposition 2.D.5 and Equation (2.14),

$$\mu = \frac{1 - 2\left(\frac{\tilde{A}-1}{\tilde{A}^2-1}\right)}{2\left(\frac{\tilde{A}-1}{\tilde{A}^2-1}\right) - 2\left(\frac{\ln\tilde{A}}{\tilde{A}^2-1}\right)} = \left(\frac{1}{2}\right)\frac{(\tilde{A}^2-1) - 2(\tilde{A}-1)}{(\tilde{A}-1) - \ln\tilde{A}} = \left(\frac{1}{2}\right)\frac{(\tilde{A}-1)^2}{(\tilde{A}-1) - \ln\tilde{A}}$$

Finally, if $k \neq 0, -1, -2$,

$$\mu = \left(\frac{k+2}{k}\right)\frac{(k+1) - k\left(\frac{\tilde{A}^{k+1}-1}{\tilde{A}^{k+1}-\tilde{A}}\right)}{(k+2)\left(\frac{\tilde{A}^{k+1}-1}{\tilde{A}^{k+1}-\tilde{A}}\right) - (k+1)\left(\frac{\tilde{A}^{k+2}-1}{\tilde{A}^{k+2}-\tilde{A}^2}\right)}$$

$$\mu = \left(\frac{k+2}{k}\right)\frac{(k+1)\left(\tilde{A}^{k+2} - \tilde{A}^2\right) - k\left(\tilde{A}^{k+2} - \tilde{A}\right)}{(k+2)\left(\tilde{A}^{k+2} - \tilde{A}\right) - (k+1)\left(\tilde{A}^{k+2} - 1\right)}$$

$$\mu = \left(\frac{k+2}{k}\right)\frac{\tilde{A}^{k+2} + k\tilde{A} - (k+1)\tilde{A}^2}{\tilde{A}^{k+2} + (k+1) - (k+2)\tilde{A}}$$

$$\mu = \left(\frac{k+2}{k}\right)\frac{\tilde{A}^2\left(\tilde{A}^k - 1\right) - k\tilde{A}\left(\tilde{A}-1\right)}{\tilde{A}\left(\tilde{A}^{k+1} - 1\right) - (k+1)\left(\tilde{A}-1\right)}$$

where we once again use Proposition 2.D.5 in Equation (2.14). ■

Proposition 2.D.7 $\lim_{\overline{A}\to+\infty} \mu = \begin{cases} \frac{k+2}{k} & , \textit{if } k > 0 \\ +\infty & , \textit{if } k < 0 \end{cases}$

Proof. Note $\lim_{\overline{A}\to+\infty} \mu = \lim_{\tilde{A}\to+\infty} \mu$ as μ is only a function of $\tilde{A} \equiv \overline{A}/\underline{A}$ (Proposition 2.D.6). Given that, it is sufficient to compute $\lim_{\tilde{A}\to+\infty} \mu$. If $k > 0$ and thus

$k + j \neq 0$ for $j \in \mathbb{N} \setminus \{0\}$, from Proposition 2.D.5 one gets

$$\lim_{\tilde{A}\to+\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \left(\frac{k}{k+j}\right) \lim_{\tilde{A}\to+\infty} \left(\frac{1-\tilde{A}^{-k-j}}{1-\tilde{A}^{-k}}\right) = \frac{k}{k+j}$$

which implies from Equation (2.14) that

$$\lim_{\tilde{A}\to+\infty} \mu = \lim_{\tilde{A}\to+\infty} \frac{1-\mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left(\underline{A}/A\right) - \mathrm{E}_a\left[(\underline{A}/A)^2\right]} = \frac{1-\frac{k}{k+1}}{\frac{k}{k+1}-\frac{k}{k+2}} = \frac{k+2}{k}$$

If $k < 0$, using Proposition 2.D.5 for $k + j \neq 0$ and $k + j = 0$, respectively,

$$\lim_{\tilde{A}\to+\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \left(\frac{k}{k+j}\right) \lim_{\tilde{A}\to+\infty} \left(\frac{\tilde{A}^k-\tilde{A}^{-j}}{\tilde{A}^k-1}\right) = 0$$
$$\lim_{\tilde{A}\to+\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \left(\lim_{\tilde{A}\to+\infty} \frac{k}{\tilde{A}^k-1}\right)\left(\lim_{\tilde{A}\to+\infty} \frac{\ln \tilde{A}}{\tilde{A}^{-k}}\right) = k\left(\lim_{\tilde{A}\to+\infty} \frac{1/\tilde{A}}{k\tilde{A}^{-k-1}}\right) = 0$$

where we apply L'Hôpital's rule in the last line, implying $\lim_{\tilde{A}\to+\infty} \mu = +\infty$ if $k < 0$ from (2.14). ■

Proposition 2.D.8 *For $k \neq 0$, $\tilde{A} \equiv \overline{A}/\underline{A}$, $\tilde{A} > 1$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu > 1$ and $k < \frac{2}{\mu-1}$.*

Proof. Let $\hat{g}$ be the density of a truncated Pareto distribution with shape parameter $k \neq 0$ defined for $A \in [\underline{A}, A_h]$, $A_h > \overline{A} > \underline{A}$, with $\hat{G}$ being the respective cumulative distribution function. It is easy to see $\tilde{g}(A) = \hat{g}(A)/\hat{G}(\overline{A})$ is the density of a truncated Pareto distribution with the same parameter $k \neq 0$ over the support $\left[\underline{A}, \overline{A}\right]$. As a consequence, we can use Proposition 2.D.4 and thus $\overline{A}$, $\overline{A} > \underline{A}$, is continuous, strictly increasing, and well defined in μ if and only if $\mu \in (1, \lim_{\overline{A}\to+\infty} \mu)$. From Proposition 2.D.6, μ is only a function of $\tilde{A} \equiv \overline{A}/\underline{A}$ (given k), implying that these features of $\overline{A}$ also hold for $\tilde{A}$. Moreover, given Proposition 2.D.7, one can rewrite the condition $\mu \in (1, \lim_{\overline{A}\to+\infty} \mu)$ as (i) $\mu > 1$ and $k < 0$ or (ii) $1 < \mu < \frac{k+2}{k}$ and $k > 0$. Note $1 < \mu < \frac{k+2}{k} \to k < \frac{2}{\mu-1}$, which is always fulfilled for $k < 0$ given $\mu > 1$. Therefore, for any $k \neq 0$, $\mu \in (1, \lim_{\overline{A}\to+\infty} \mu)$ holds if and only if $\mu > 1$ and $k < \frac{2}{\mu-1}$. ■

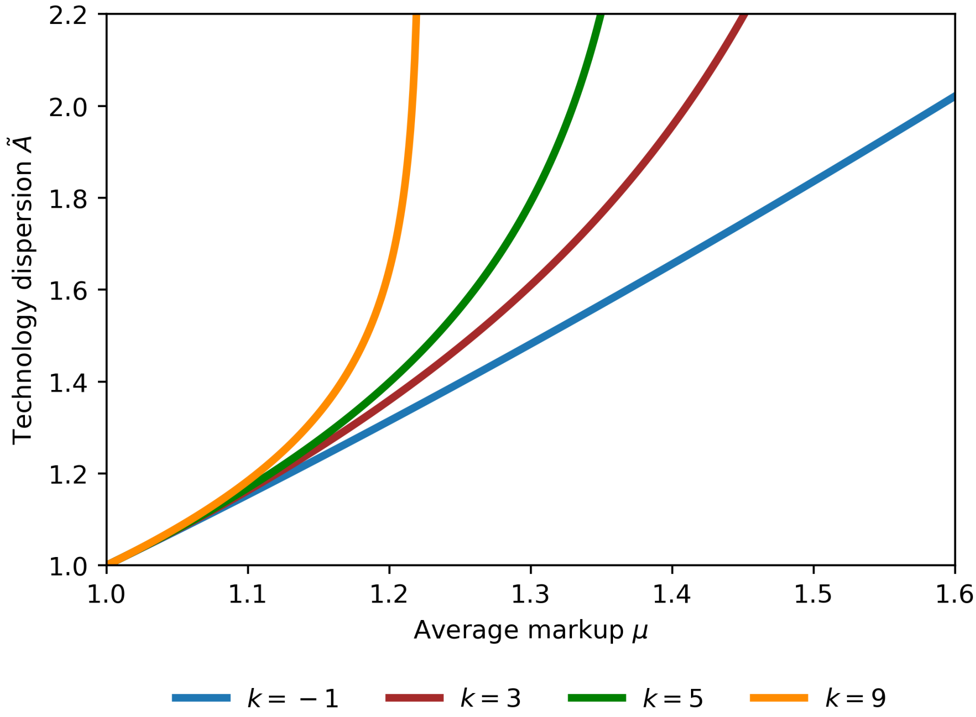


Figure 2.D.1 – Technology dispersion vs. average markup.

Figure 2.D.1 illustrates the results of Proposition 2.D.8, plotting $\tilde{A}$ against μ for truncated Pareto distributions with $k = 3, 5, 9$ and an Uniform distribution $(k = -1)$.

From Proposition 2.D.8, given $\mu > 1$, a solution for the calibration algorithm exists for any strictly negative shape parameter k. Hence, it is feasible even for highly negative values of k. However, an important inquiry arises regarding the behavior of the estimated Ω under such extreme conditions. This question is addressed in the concluding proposition of this section:

Proposition 2.D.9 $\lim_{k\to-\infty} \Omega = 1$.

Proof. Initially, use Proposition 2.D.5 to get that, for $j \in \mathbb{N} \setminus \{0\}$,

$$\lim_{k\to-\infty} \mathrm{E}_a\left((\underline{A}/A)^j\right) = \left[\lim_{k\to-\infty}\left(\frac{k}{k+j}\right)\right]\left[\lim_{k\to-\infty}\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right)\right] = \frac{0-1}{0-\tilde{A}^j} = \tilde{A}^{-j}$$

As a result, from Equation (2.11) holding exactly,

$$\lim_{k\to-\infty}\Omega=\lim_{k\to-\infty}\frac{\tilde{A}^{-1}\left[1-\mathrm{E}_a\left(\underline{A}/A\right)\right]}{\mathrm{E}_a\left(\underline{A}/A\right)-\mathrm{E}_a\left[(\underline{A}/A)^2\right]}=\frac{\tilde{A}^{-1}\left(1-\tilde{A}^{-1}\right)}{\tilde{A}^{-1}-\tilde{A}^{-2}}=\frac{\tilde{A}-1}{\tilde{A}-1}=1$$

where we use $\lim_{k\to-\infty}\mathrm{E}_a\left((\underline{A}/A)^j\right)=\tilde{A}^{-j}$ for $j=1,2$. ■

2.E Robustness of misallocation estimates to markup level

We want to prove that, for $k=-1$ and $\mu>1$, estimates of allocative efficiency *growth* are highly robust to (i) the level of LS and (ii) the choice of α. More precisely, let us show that $\Delta\ln\Omega$ is highly robust to the level of $\mu=\frac{1-\alpha}{LS}$. Initially, use Equation (2.12) to get

$$\Omega=\mu/\tilde{A}\to\Delta\ln\Omega=\Delta\ln\mu-\Delta\ln\tilde{A} \tag{2.E.1}$$

implying $\Delta\ln\Omega$ is independent of the level of μ if $\ln\tilde{A}=\gamma_0+\gamma_1\ln\mu$, when (2.E.1) becomes $\Delta\ln\Omega=(1-\gamma_1)\Delta\ln\mu$. We use this idea to seek the k that yields a nearly independent $\Delta\ln\Omega$ with respect to the level of μ by finding the k that yields the best fit for the regression $\ln\tilde{A}(\mu;k)=\gamma_0+\gamma_1\ln\mu+\varepsilon$, where $\tilde{A}(\mu;k)$ is implicitly defined in (2.14). We consider 100 observations for μ between 1 and μ_h, for μ_h equal to 1.05, 1.15, 1.3, 1.5, 3, 5 or 10.[35] For each μ_h and thus each set of observations of μ, choosing a k, we obtain the respective 100 observations of $\tilde{A}$ numerically from (2.14), estimating γ_0 and γ_1 using Ordinary Least Squares (OLS). Hence, for each μ_h, we can obtain the sum of squared residuals (SSR) associated with any $k<\frac{2}{\mu_h-1}$, $k\neq0$, allowing the numerical search of the robust k that minimizes the SSR.

The robust k, denoted by k^*, and the associated (centered) R^2 for each μ_h are shown respectively in the second and third columns of Table 2.E.1. As can be seen, the R^2 is always very high, with k^* being close to -1, particularly for smaller μ_h. In any case, even when k^* is not so close to -1, the R^2 associated with $k=-1$

[35] Empirical evidence from firm-level data suggests $\mu=10$ is high enough (Loecker; Eeckhout; Unger, 2020; Traina, 2018; Baqaee; Farhi, 2020; Loecker; Eeckhout, 2018). For instance, Loecker and Eeckhout (2018) find that global cost-weighted average markups are always below 1.6. The sales-weighted average, which is typically a higher measure (Loecker; Eeckhout; Unger, 2020), remains below 2 for global regions (Europe, North America, South America, Asia, Oceania, and Africa) and under 4 for a selection of 40 countries shown in their Appendix A.

is still very high (fourth column of Table 2.E.1), indicating the estimate of $\Delta \ln \Omega$ under $k = -1$ is also highly robust to the level of μ.

Another way to see that is evaluating $\frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu}$. Given μ_h and the correspondent set of observations $\{\mu_i\}$, it is possible to show $\frac{\partial \ln \tilde{A}(\mu_i;-1)}{\partial \ln \mu} \in \left[\frac{\partial \ln \tilde{A}(\mu_h;-1)}{\partial \ln \mu}, 1.5\right) \subset (1, 1.5)$. These derivative ranges for each considered μ_h are shown in the last column of Table 2.E.1.[36] The ranges are always very narrow, especially for low μ_h, confirming that the estimate of $\Delta \ln \Omega$ under $k = -1$ is highly robust to the choice of α and the level of LS.[37]

Table 2.E.1 – Robustness evaluation for different μ_h

μ_h	$k = k^*$		$k = -1$	
	k^*	R^2	R^2	$\frac{\partial \ln \tilde{A}}{\partial \ln \mu}$ range
1.05	-0.9949	1.000000	1.000000	[1.4999,1.5)
1.15	-0.9860	1.000000	1.000000	[1.4993,1.5)
1.3	-0.9738	1.000000	1.000000	[1.4974,1.5)
1.5	-0.9596	1.000000	1.000000	[1.4939,1.5)
3	-0.8931	1.000000	0.999986	[1.4590,1.5)
5	-0.8480	0.999998	0.999944	[1.4205,1.5)
10	-0.7940	0.999994	0.999817	[1.3632,1.5)

This derivative assessment relied on some results. We prove them now. To start, note μ is a function of $\tilde{A}$ from Equation (2.14), which for the Uniform distribution is equivalent to

$$\mu\left(\tilde{A}; -1\right) = \frac{\left(\tilde{A} - 1\right) - \ln \tilde{A}}{\ln \tilde{A} - \left(\frac{\tilde{A}-1}{\tilde{A}}\right)} \tag{2.E.2}$$

[36] To get $\frac{\partial \ln \tilde{A}(\mu_h;-1)}{\partial \ln \mu}$, we first obtain $\tilde{A}(\mu_h; -1)$ numerically from (2.E.2) and then compute (2.E.3) to get $\frac{\partial \ln \mu\left(\tilde{A}(\mu_h;-1);-1\right)}{\partial \ln \tilde{A}}$. Finally, we use Proposition 2.E.4 shown below and compute $\frac{\partial \ln \tilde{A}(\mu_h;-1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu_h;-1);-1\right)}{\partial \ln \tilde{A}}\right]^{-1}$.

[37] Up to a first-order Taylor approximation around $\mu^* > 1$, $\ln \tilde{A}(\mu; -1) \approx \ln \tilde{A}(\mu^*; -1) + \frac{\partial \ln \tilde{A}(\mu^*;-1)}{\partial \ln \mu}(\ln \mu - \ln \mu^*)$. However, for $\mu^* \to 1^+$, $\tilde{A}(\mu; -1) \to 1$ (Propositions 2.D.3 and 2.D.4) and $\frac{\partial \ln \tilde{A}(\mu^*;-1)}{\partial \ln \mu} \to 1.5$ (Proposition 2.E.4 shown below). Consequently, for low average markup, $\ln \tilde{A}(\mu; -1) \approx 1.5 \ln \mu \to \tilde{A}(\mu; -1) \approx \mu^{1.5}$ and thus, from (2.E.1), $\Delta \ln \Omega \approx -0.5\Delta \ln \mu = 0.5\Delta \ln LS$.

due to Proposition 2.D.6. As a consequence,

$$\frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \frac{\tilde{A}-1}{\left(\tilde{A}-1\right)-\ln \tilde{A}} - \frac{1-\frac{\tilde{A}}{\tilde{A}^2}}{\ln \tilde{A}-\left(\frac{\tilde{A}-1}{\tilde{A}}\right)}$$

$$\frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \frac{\tilde{A}-1}{\left(\tilde{A}-1\right)-\ln \tilde{A}} - \frac{\tilde{A}-1}{\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)}$$

$$\frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \frac{\left(\tilde{A}-1\right)\left[\left(\tilde{A}+1\right)\ln \tilde{A}-2\left(\tilde{A}-1\right)\right]}{\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]} \tag{2.E.3}$$

$$\frac{\partial^2 \ln \mu\left(\tilde{A};-1\right)}{\partial^2 \ln \tilde{A}} = \frac{\tilde{A}\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]-\left(\tilde{A}-1\right)^2}{\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]^2} - \frac{\tilde{A}\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]-\left(\tilde{A}-1\right)\tilde{A}\ln \tilde{A}}{\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]^2}$$

$$\frac{\partial^2 \ln \mu\left(\tilde{A};-1\right)}{\partial^2 \ln \tilde{A}} = \frac{\left(\tilde{A}-1\right)-\tilde{A}\ln \tilde{A}}{\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]^2} - \frac{\tilde{A}\left[\ln \tilde{A}-\left(\tilde{A}-1\right)\right]}{\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]^2}$$

$$\frac{\partial^2 \ln \mu\left(\tilde{A};-1\right)}{\partial^2 \ln \tilde{A}} = \frac{\tilde{A}\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]^3-\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]^3}{\left[\left(\tilde{A}-1\right)-\ln \tilde{A}\right]^2\left[\tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)\right]^2} \tag{2.E.4}$$

Proposition 2.E.1 *For* $\tilde{A} > 1$, $\frac{\partial^2 \ln \mu\left(\tilde{A};-1\right)}{\partial^2 \ln \tilde{A}} > 0$.

Proof. Let $\tilde{A} > 1$. In this case, (i) for $g(\tilde{A}) \equiv \left(\tilde{A}-1\right)-\ln \tilde{A}$, $g'(\tilde{A}) = \frac{\tilde{A}-1}{\tilde{A}} > 0 \xrightarrow{g(1)=0} g(\tilde{A}) > 0$, and (ii) for $h(\tilde{A}) \equiv \tilde{A}\ln \tilde{A}-\left(\tilde{A}-1\right)$, $h'(\tilde{A}) = \ln \tilde{A} > 0 \xrightarrow{h(1)=0} h(\tilde{A}) > 0$. Consequently, the denominator of (2.E.4) is strictly positive. Hence, it is sufficient to show $f(\tilde{A}) \equiv \sqrt[3]{\tilde{A}}\left[(\tilde{A}-1)-\ln \tilde{A}\right]-\left[\tilde{A}\ln \tilde{A}-(\tilde{A}-1)\right] > 0$. Note $f(1) = 0$ and

$$f'(\tilde{A}) = \frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left[(\tilde{A}-1)-\ln \tilde{A}\right]+\sqrt[3]{\tilde{A}}\left(\frac{\tilde{A}-1}{\tilde{A}}\right)-\ln \tilde{A}$$

$$f'(\tilde{A}) = \frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left[4(\tilde{A}-1)-\ln \tilde{A}\right]-\ln \tilde{A} \rightarrow f'(1) = 0$$

$$
\begin{aligned}
f''(\tilde{A}) =& -2\frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[4(\tilde{A}-1)-\ln\tilde{A}\right]+\frac{\sqrt[3]{\tilde{A}}}{3\tilde{A}}\left(\frac{4\tilde{A}-1}{\tilde{A}}\right)-\frac{1}{\tilde{A}}\\
f''(\tilde{A}) =& \frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[-8(\tilde{A}-1)+2\ln\tilde{A}+12\tilde{A}-3-9\tilde{A}^{2/3}\right]\\
f''(\tilde{A}) =& \frac{\sqrt[3]{\tilde{A}}}{9\tilde{A}^2}\left[4\tilde{A}+5+2\ln\tilde{A}-9\tilde{A}^{2/3}\right]
\end{aligned}
$$

$$
\begin{aligned}
\tilde{f}(\tilde{A}) \equiv& 4\tilde{A}+5+2\ln\tilde{A}-9\tilde{A}^{2/3}\to\tilde{f}(1)=0\\
\tilde{f}'(\tilde{A}) =& 4+\frac{2}{\tilde{A}}-\frac{6}{\sqrt[3]{\tilde{A}}}\to\tilde{f}'(1)=0\\
\tilde{f}''(\tilde{A}) =& -\frac{2}{\tilde{A}^2}+\frac{2}{\tilde{A}\sqrt[3]{\tilde{A}}}=\frac{2}{\tilde{A}^2}\left(\tilde{A}^{2/3}-1\right)\to\tilde{f}''(\tilde{A})>0 \text{ for } \tilde{A}>1
\end{aligned}
$$

Therefore, for $\tilde{A}>1$, $\tilde{f}''(\tilde{A})>0 \xrightarrow{\tilde{f}'(1)=0} \tilde{f}'(\tilde{A})>0 \xrightarrow{\tilde{f}(1)=0} \tilde{f}(\tilde{A})>0 \to f''(\tilde{A})>0 \xrightarrow{f'(1)=0} f'(\tilde{A})>0 \xrightarrow{f(1)=0} f(\tilde{A})>0$. ■

Proposition 2.E.2 $\lim_{\tilde{A}\to+\infty}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}}=1$.

Proof. From Equation (2.E.3),

$$
\begin{aligned}
\lim_{\tilde{A}\to+\infty}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} =& \frac{1}{1-\lim_{\tilde{A}\to+\infty}\frac{\ln\tilde{A}}{\tilde{A}-1}}-\frac{1}{\lim_{\tilde{A}\to+\infty}\frac{\tilde{A}\ln\tilde{A}}{\tilde{A}-1}-1}\\
\lim_{\tilde{A}\to+\infty}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}} =& \frac{1}{1-\lim_{\tilde{A}\to+\infty}\frac{1}{\tilde{A}}}-\frac{1}{\lim_{\tilde{A}\to+\infty}\left(\ln\tilde{A}+1\right)-1}=1-0=1
\end{aligned}
$$

where we apply L'Hôpital's rule in the second line. ■

Proposition 2.E.3 $\lim_{\tilde{A}\to1^+}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}}=\frac{2}{3}$.

Proof. From Equation (2.E.3),

$$
\lim_{\tilde{A}\to1^+}\frac{\partial\ln\mu\left(\tilde{A};-1\right)}{\partial\ln\tilde{A}}=\lim_{\tilde{A}\to1^+}\frac{\left(\tilde{A}+1\right)\ln\tilde{A}-2\left(\tilde{A}-1\right)+\left(\tilde{A}-1\right)\left(\ln\tilde{A}+\frac{\tilde{A}+1}{\tilde{A}}-2\right)}{\left(\frac{\tilde{A}-1}{\tilde{A}}\right)\left[\tilde{A}\ln\tilde{A}-\left(\tilde{A}-1\right)\right]+\left[\left(\tilde{A}-1\right)-\ln\tilde{A}\right]\ln\tilde{A}}
$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \tilde{A}\frac{2\left(\ln \tilde{A} + \frac{\tilde{A}+1}{\tilde{A}} - 2\right) + \left(\frac{\tilde{A}-1}{\tilde{A}}\right)^2}{\ln \tilde{A} - \frac{\tilde{A}-1}{\tilde{A}} + 2\left(\tilde{A}-1\right)\ln \tilde{A} + \left(\tilde{A}-1\right) - \ln \tilde{A}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} = \lim_{\tilde{A}\to 1^+} \frac{2\left(\frac{1}{\tilde{A}} - \frac{1}{\tilde{A}^2}\right) + 2\left(\frac{\tilde{A}-1}{\tilde{A}}\right)\left(\frac{1}{\tilde{A}^2}\right)}{-\frac{1}{\tilde{A}^2} + 2\ln \tilde{A} + 2\left(\frac{\tilde{A}-1}{\tilde{A}}\right) + 1}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} =2 \lim_{\tilde{A}\to 1^+} \frac{\left(\frac{\tilde{A}-1}{\tilde{A}^2}\right)\left(\frac{\tilde{A}+1}{\tilde{A}}\right)}{2\ln \tilde{A} + 3 - \frac{2}{\tilde{A}} - \frac{1}{\tilde{A}^2}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} =2 \lim_{\tilde{A}\to 1^+} \frac{\tilde{A}^2 - 1}{2\tilde{A}^3 \ln \tilde{A} + 3\tilde{A}^3 - 2\tilde{A}^2 - \tilde{A}}$$

$$\lim_{\tilde{A}\to 1^+} \frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}} =2 \lim_{\tilde{A}\to 1^+} \frac{2\tilde{A}}{6\tilde{A}^2 \ln \tilde{A} + 2\tilde{A}^2 + 9\tilde{A}^2 - 4\tilde{A} - 1} = 2\left(\frac{2}{6}\right) = \frac{2}{3}$$

where we apply L'Hôpital's rule in the first, second, third, and sixth lines. ■

Proposition 2.E.4 $\frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu;-1);-1\right)}{\partial \ln \tilde{A}}\right]^{-1}$ *is continuous and strictly decreasing in* $\mu > 1$*, with* $\lim_{\mu\to+\infty} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = 1$ *and* $\lim_{\mu\to 1^+} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = 1.5$.

Proof. From Proposition 2.D.8, $\tilde{A}(\mu;-1)$, which is implicitly defined in (2.E.2), is continuous, strictly increasing, and well defined for $\mu > 1$. Moreover,

$$1 = \frac{\partial \ln \tilde{A}}{\partial \ln \tilde{A}} = \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu}\frac{\partial \ln \mu\left(\tilde{A};-1\right)}{\partial \ln \tilde{A}}$$

$$\therefore \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \left[\frac{\partial \ln \mu\left(\tilde{A}(\mu;-1);-1\right)}{\partial \ln \tilde{A}}\right]^{-1}$$

Hence, since $\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}$ is continuous and strictly increasing in $\tilde{A} > 1$ (Proposition 2.E.1), $\frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu}$ is continuous and strictly decreasing in $\mu > 1$. Furthermore, using Propositions 2.E.2 and 2.E.3,

$$\lim_{\mu\to+\infty} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \lim_{\tilde{A}\to+\infty} \frac{1}{\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}} = 1$$

$$\lim_{\mu\to 1^+} \frac{\partial \ln \tilde{A}(\mu;-1)}{\partial \ln \mu} = \lim_{\tilde{A}\to 1^+} \frac{1}{\frac{\partial \ln \mu(\tilde{A};-1)}{\partial \ln \tilde{A}}} = 1.5$$

since (i) $\tilde{A}(\mu;-1) \to +\infty$ if and only if $\mu \to +\infty$ (Propositions 2.D.7 and 2.D.8) and (ii) $\tilde{A}(\mu;-1) \to 1^+$ if and only if $\mu \to 1^+$ (Propositions 2.D.3 and 2.D.4). ■

2.F Model extensions

We develop two model extensions. In both cases, we consider a discrete number of firms as in Section 2.2.1, but they can easily be adapted to a continuum of firms following the approach of Section 2.2.2.

2.F.1 Beyond Cobb-Douglas production functions

In this section, we generalize the model by considering an arbitrary well-behaved production function with M factors of production, provided it exhibits (i) constant returns to scale and (ii) Hicks-neutral productivity shifter. Formally, the production of each firm $i \in \{1, 2, ..., N\}$ is now given by

$$Y_i = A_i f(F_{1i}, ..., F_{Mi}) \tag{2.2'}$$

where $A_i > 0$ and $F_{ji} \geq 0$ is the quantity of factor j used by firm i. The function f is (i) homogeneous of degree 1, (ii) differentiable, and (iii) well-behaved in the sense that the firm's cost minimization problem has a unique interior solution for strictly greater than zero rental prices.

Before discussing the derivation of the model, note that if each firm i uses its inputs optimally, taking the same inputs' rental prices as given, the FOC of its cost minimization problem implies that, for any inputs r and k,

$$\frac{w_r}{MPF_{ri}} = \frac{w_k}{MPF_{ki}} \leftrightarrow \frac{w_r}{w_k} = \frac{A_i f_r(F_{1i}, ..., F_{Mi})}{A_i f_k(F_{1i}, ..., F_{Mi})} = \frac{f_r(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})}{f_k(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})} \tag{2.A.3'}$$

where $w_r > 0$ is the rental price of factor r, MPF_{ri} is firm i marginal product of factor r, $f_r(F_{1i}, ..., F_{Mi}) \equiv \frac{\partial f(F_{1i}, ..., F_{Mi})}{\partial F_{ri}}$, and in the last part we use that f_r is homogeneous of degree 0 as f is homogeneous of degree 1. Setting $r = 1, 2, ..., k-1, k+1, ..., M$ on (2.A.3'), we would get a system of $M-1$ equations in the $M-1$ unknowns $F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i}$. Since the cost minimization problem has a unique interior solution by assumption, the FOCs are satisfied in the optimal, and thus, there is a unique solution for this system. Hence, since all firms face the same problem, they choose the same relative quantities of factors, that is, $F_{ri}/F_{ki} = F_{rj}/F_{kj}$ for any firms i and j and factors r and k.

2.F.1.1 Derivation of the model

Given its similarity to the baseline model, we do not discuss the derivation in detail, emphasizing solely the main differences from the Cobb-Douglas case.

Environment and technology. The only difference is the firms' production function, which is now represented by (2.2') instead of (2.2). Aggregate output Y continues to be given by (2.1).

Market competition and optimal decision. As the marginal cost remains invariant to output, Equations (2.3) and (2.4) of the Cobb-Douglas case persist, albeit with a distinct marginal cost function. The SOC is the same (2.A.1) and thus holds.

Equilibrium allocation. Since firms use inputs optimally, $F_{ri}/F_{ki} = F_{rj}/F_{kj}$ and $MC_i = \frac{w_k}{MPF_{ki}} = \frac{w_k}{A_i f_k(1, F_{2i}/F_{1i}, ..., F_{Mi}/F_{1i})}$ for any firms i and j and factors r and k. As a consequence, $\frac{MC_i}{MC_j} = \frac{A_j}{A_i}$ as before, implying Equations (2.5), (2.A.2), (2.6), and (2.7) are still valid.

Aggregate productivity and misallocation. Naturally, the aggregate production function is not (2.8) anymore. However, using (2.1) and (2.2'), one easily obtains the new expression:

$$Y = \sum_{i=1}^{N} A_i f(F_{1i}, ..., F_{Mi}) = \overline{A}\Omega f(F_1, ..., F_M) \tag{2.8'}$$

where $F_j \equiv \sum_{i=1}^{N} F_{ji}$, and $\Omega \equiv \sum_{i=1}^{N} \frac{f(F_{1i}, ..., F_{Mi})}{f(F_1, ..., F_M)} \left(A_i/\overline{A}\right)$.

Since each firm uses its inputs optimally, taking the same inputs' rental prices as given, every active firm chooses the same relative quantities of factors, implying

$$\theta_{rj} \equiv \frac{F_{rj}}{\sum_{i=1}^{N} F_{ri}} = \frac{F_{kj}\frac{F_{rj}}{F_{kj}}}{\sum_{i=1}^{N} F_{ki}\frac{F_{ri}}{F_{ki}}} = \frac{F_{kj}\frac{F_{rj}}{F_{kj}}}{\sum_{i=1}^{N} F_{ki}\frac{F_{rj}}{F_{kj}}} = \frac{F_{kj}}{\sum_{i=1}^{N} F_{ki}} \equiv \theta_{kj} \tag{2.A.4'}$$

which allows us to define $\theta_j \equiv \theta_{1j} = \theta_{2j} = ... = \theta_{Mj}$ for each firm j and factors r and k. As a result, from Equations (2.2') and (2.8'),

$$s(A_i) = \frac{Y_i}{Y} = \frac{A_i f(F_{1i}, ..., F_{Mi})}{\overline{A}\Omega f(F_1, ..., F_M)} = \frac{A_i f(\theta_i F_1, ..., \theta_i F_M)}{\overline{A}\Omega f(F_1, ..., F_M)} = \frac{A_i \theta_i}{\overline{A}\Omega} \rightarrow \theta_i = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{2.9'}$$

which is exactly equal to (2.9). Consequently, the aggregate TFP $\overline{A}\Omega$ continue to be the quantity-weighted harmonic mean of firms' productivity shown in (2.10), with allocative efficiency Ω still given by Equation (2.11) and thus showing the same properties as before.

Average markup. Given that the main results from the baseline model continue to be valid, it is easy to see that cost-weighted average of firm-level markups $\mu \equiv \sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{M} F_{ki} w_k}{\sum_{k=1}^{M} F_k w_k} \right) \mu_i = \sum_{i=1}^{N} \theta_i \mu_i$ is still given by Equation (2.12).

2.F.1.2 Quantification strategy

The target moments $\overline{A}\Omega$ and μ can be computed in the model using the same expressions of the Cobb-Douglas case. Consequently, given data on such moments, we can empirically implement this generalized model following exactly the strategy described in Section 2.3.3. The difference between models appears only when computing those target moments from data.

Analogously to the baseline case, we get $\overline{A}\Omega$ and μ using standard macroeconomic data and the parameters of the production function. On the one hand, TFP continues to be backed out as a residual in the aggregate output function, which is now given by Equation (2.8'): $\overline{A}\Omega = \frac{Y}{f(F_1,...,F_M)}$. On the other hand, note for any firm i, $\mu_i = \frac{Y_i p}{H_i w} \frac{H_i MPH_i}{Y_i} = \frac{\alpha_H}{LS_i}$, where $\alpha_H \equiv \frac{H_i MPH_i}{Y_i} = \frac{H_i f_1(F_{1i}, F_{2i},...,F_{Mi})}{f(F_{1i}, F_{2i},...,F_{Mi})} = \frac{f_1(1, F_{2i}/H_i,...,F_{Mi}/H_i)}{f(1, F_{2i}/H_i,...,F_{Mi}/H_i)} = \frac{f_1(1, F_2/H,...,F_M/H)}{f(1, F_2/H,...,F_M/H)}$ is the elasticity of the production function to human capital, which we assume is factor 1. As a result, $\mu = \sum_{i=1}^{N} \theta_i \mu_i = \sum_{i=1}^{N} \left(\frac{H_i}{H} \right) \left[\alpha_H \frac{Y_i p}{H_i w} \right] = \alpha_H \frac{p \sum_{i=1}^{N} Y_i}{Hw} = \frac{\alpha_H}{LS}$. Note computing $\alpha_H = \frac{f_1(1, F_2/H,...,F_M/H)}{f(1, F_2/H,...,F_M/H)}$ requires, in general, the relative aggregate quantities of factors, while in the special Cobb-Douglas case it does not as α_H would be simply the share parameter associated with human capital. Indeed, for the baseline model, in which $f(H, K) = K^{\alpha} H^{1-\alpha}$, $\alpha_H = \frac{f_1(1, K/H)}{f(1, K/H)} = \frac{(1-\alpha)(K/H)^{\alpha}(H/H)^{-\alpha}}{(H/H)^{\alpha}(H/H)^{1-\alpha}} = 1 - \alpha$, implying

$\mu = \frac{1-\alpha}{LS}$, which is exactly the result shown in Section 2.3.2. Thus, differently from the Cobb-Douglas case, α_H is not generally time invariant even if the production function parameters are kept constant, meaning the average markup μ growth may differ from the labor share growth.

2.F.1.3 Calibration of the production function

In Section 2.4.1, we calibrate the baseline Cobb-Douglas production function for the US using cost share data. A similar approach can be employed here, since

$$\frac{Hw}{\sum_{k=1}^{M} F_k w_k} = \frac{\theta_i H w}{\sum_{k=1}^{M} \theta_i F_k w_k} = \frac{H_i}{\sum_{k=1}^{M} F_{ki} \frac{w_k}{w}} = \frac{H_i}{\sum_{k=1}^{M} F_{ki} \frac{f_k(1,F_{2i}/H_i,...,F_{Mi}/H_i)}{f_1(1,F_{2i}/H_i,...,F_{Mi}/H_i)}}$$
$$\frac{Hw}{\sum_{k=1}^{M} F_k w_k} = \frac{f_1(1,F_{2i}/H_i,...,F_{Mi}/H_i)}{\sum_{k=1}^{M} \frac{F_{ki}}{H_i} f_k(1,F_{2i}/H_i,...,F_{Mi}/H_i)} = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)} = \alpha_H \tag{2.A.5'}$$

where we use, in the first line, (2.A.3') with $F_{1i} = H_i$ and, in the last line, the Euler's Theorem for Homogeneous Functions and that, in the optimal, $F_{ki}/H_i = F_k/H$ for any factor k.[38] As a result, one can calibrate the production function by finding the parameters of f that make $\alpha_H = \frac{f_1(1,F_2/H,...,F_M/H)}{f(1,F_2/H,...,F_M/H)}$ closest (in some sense) to labor cost share data. This calibration could also include cost share data for other factors of production as $\alpha_r = \frac{f_r(F_1/F_r,...,F_M/F_r)}{f(F_1/F_r,...,F_M/F_r)} = \frac{F_r w_r}{\sum_{k=1}^{M} F_k w_k}$ for any factor r. However, one may use at most $M-1$ factors' cost share data in the calibration since share data from the remaining factor would not be informative given that $\sum_{r=1}^{M} \alpha_r = 1$.

To illustrate this calibration procedure, let us consider the CES production function

$$Y_i = A_i \left[\alpha K_i^{\frac{\sigma-1}{\sigma}} + (1-\alpha) H_i^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{2.F.1}$$

where $\sigma > 0$ is the elasticity of substitution and $\alpha \in (0,1)$. In this case,

$$\alpha_H = \frac{f_1(1,K/H)}{f(1,K/H)} = \frac{f_1(H,K)}{f(1,K/H)} = \frac{\left[\alpha K^{\frac{\sigma-1}{\sigma}} + (1-\alpha) H^{\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} (1-\alpha) H^{\frac{-(\sigma-1)/\sigma}{\sigma-1}}}{\left[\alpha (K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{\sigma}{\sigma-1}}}$$

[38] As expected, Equation (2.A.5') is consistent with (2.A.5), since in the baseline case $1-\alpha = \alpha_H = \frac{Hw}{Kr+Hw}$.

$$\alpha_H = \frac{(1-\alpha)\left[\alpha(K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{1}{\sigma-1}}}{\left[\alpha(K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\right]^{\frac{\sigma}{\sigma-1}}} = \frac{1-\alpha}{\alpha(K/H)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)} \tag{2.F.2}$$

with $\alpha_H \to 1-\alpha$ when $\sigma \to 1$, which is expected as the CES function (2.F.1) converges to the baseline Cobb-Douglas function (2.2) in this limit case. Hence, given data on K and H and the parameters α and σ, we can compute α_H from (2.F.2), allowing us to gauge α and σ by matching α_H to labor cost share data.

2.F.2 Firm-specific wedges

We consider the same setup of the baseline model but add one wedge for each potential entrant firm, including a firm-specific tax rate over revenue $\tau_i = \tau(A_i) \in (-\infty, 1)$

2.F.2.1 Derivation of the model

As in Appendix 2.F.1.1, we do not provide a complete derivation. Instead, we just highlight the key differences from the baseline model.

Environment and technology. The environment and technology remain unchanged, with Equations (2.1) and (2.2) still holding.

Market competition and optimal decision. Since now there is a revenue tax, each firm $i \in \{1, 2, ..., N\}$ solves a slightly different profit maximization problem:

$$\begin{aligned} \max_{Y_i} \quad & p\,(1-\tau(A_i))\,Y_i - wH_i - rK_i = [p\,(1-\tau(A_i)) - MC_i]\,Y_i \\ \text{s.t.} \quad & w > 0, r > 0, p = p(Y), Y_j \geq 0 \;\forall j \in \{1, 2, ..., N\} \setminus \{i\} \end{aligned} \tag{2.3''}$$

The FOC of this optimization problem is

$$p\,(1-\tau(A_i)) = MC_i \frac{\eta_i}{\eta_i - 1} \tag{2.4''}$$

while the SOC is still (2.A.1).

Equilibrium allocation. Equation (2.5) is not valid anymore, but one can obtain an analogous expression using (2.4”) for any active firms i and j:

$$\frac{MC_i}{1-\tau(A_i)}(\eta - s_i)^{-1} = \frac{MC_j}{1-\tau(A_j)}(\eta - s_j)^{-1}$$

$$s_i - s_j = \left[1 - \frac{MC_i/(1-\tau(A_i))}{MC_j/(1-\tau(A_j))}\right](\eta - s_j) = \left(1 - \frac{B_j}{B_i}\right)(\eta - s_j) \tag{2.5”}$$

where $B_i \equiv A_i\,(1-\tau(A_i))$. In the following, let $\underline{B} \equiv \min_i\{B_i\}$ and $\overline{B} \equiv \max_i\{B_i\}$, with $0 < \underline{B} < \overline{B} < +\infty$. We again discard any equilibrium in which a non-active firm has a lower marginal cost than an active firm, considering here the adjusted marginal cost $\frac{MC_i}{1-\tau(A_i)} = \left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{B_i}$. Hence, analogously to the baseline model, in the unique refined equilibrium, there exists a firm with adjusted productivity $\underline{B}$ serving as the cutoff for active firms, such that firm i is active if and only if $B_i \geq \underline{B}$.[39]

In this context, following the same steps of the baseline model but using (2.5”) instead of (2.5), one can easily show that

$$s(B_i) = \frac{1}{N_a \mathrm{E}_a\,(B_i/B)} + \eta\left(1 - \frac{1}{\mathrm{E}_a\,(B_i/B)}\right) \tag{2.A.2”}$$

where now $\mathrm{E}_a\,(h(A)) \equiv \mathrm{E}\,(h(A)|B \geq \underline{B})$.

We suppose again inefficient technologies are common knowledge, with $\underline{A} \to 0$ and $N \to \infty$. Furthermore, although firms may receive subsidies ($\tau_i < 0$), we assume these subsidies do not vanish the dispersion in B in the sense that $(A_i \to 0) \to (B_i \to 0)$. As a result, low-productivity firms will not be active, and thus, the market share $s(\underline{B})$ should be relatively low. Assuming it is approximately null and using (2.5”) and (2.A.2”), one can see that

$$s(B_i) \approx \eta\,(1 - \underline{B}/B_i) \tag{2.6”}$$

$$\eta \approx \frac{1}{N_a\,[1 - \mathrm{E}_a\,(\underline{B}/B)]} \tag{2.7”}$$

Aggregate productivity and misallocation. Equations (2.8), (2.A.3), (2.A.4), (2.A.5), (2.9), and (2.10) remain valid, with $s(A_i)$ replaced by $s(B_i)$. However, allocative

[39] Similarly to the baseline discrete model, for the sake of clarity and convenience, we rule out situations where not all firms with productivity $\underline{B}$ are active. As before, this results in just a low loss of generality.

efficiency Ω is not given anymore by (2.11). To obtain the new expression, just plug (2.6") and (2.7") into (2.10):

$$\frac{1}{\overline{A}\Omega} \approx \frac{\mathrm{E}_a(1/A) - \mathrm{E}_a\left[\underline{B}/(A \times B)\right]}{1 - \mathrm{E}_a\left(\underline{B}/B\right)} = \frac{\mathrm{E}_a(\underline{A}/A) - \mathrm{E}_a\left[(\underline{A} \times \underline{B})/(A \times B)\right]}{\underline{A}\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]}$$

$$\Omega \approx \frac{(\underline{A}/\overline{A})\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]}{\mathrm{E}_a(\underline{A}/A) - \mathrm{E}_a\left[(\underline{A} \times \underline{B})/(A \times B)\right]} = \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{B}/B)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{B}/B)\right]} \tag{2.11"}$$

Average markup. The cost-weighted average markup μ no longer follows Equation (2.12). However, given (i) (2.9) with $s(A_i)$ replaced by $s(B_i)$, (ii) (2.4"), (iii) that the active firm with adjusted productivity $\underline{B}$ has markup approximately equal to one, and (iv) the Cobb-Douglas marginal cost function, it is easy to show that

$$\mu = \sum_{i=1}^{N} \theta_i \mu_i \approx \sum_{i=1}^{N} \left[\overline{A}\Omega \frac{s(B_i)}{A_i}\right] \left[\frac{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{\underline{B}}}{\left(\frac{r}{\alpha}\right)^{\alpha}\left(\frac{w}{1-\alpha}\right)^{1-\alpha}\frac{1}{B_i}}\right] = \frac{\overline{A}\Omega}{\underline{B}}\left(1 - \tilde{\tau}\right) \tag{2.12"}$$

where $\tilde{\tau} \equiv \sum_{i=1}^{N} s(B_i)\tau(A_i)$ is the sales-weighted average tax rate.

Average tax rate. We have already derived all equations analogous to the baseline model.[40] However, this model has an additional equation, for the sales-weighted average tax rate $\tilde{\tau}$:

$$\tilde{\tau} \approx \frac{\mathrm{E}_a\left[\tau(A)(1 - \underline{B}/B)\right]}{\mathrm{E}_a\left(1 - \underline{B}/B\right)} \tag{2.F.3}$$

which can be easily obtained by plugging (2.6") and (2.7") into $\tilde{\tau} \equiv \sum_{i=1}^{N} s(B_i)\tau(A_i)$.

2.F.2.2 Remarks on the quantification strategy

We briefly discuss the quantification of this model under two scenarios. First, assuming the function τ is known, one can essentially follow the strategy of Section 2.3, but using (2.11") and (2.12") instead of (2.11) and (2.12). Second, suppose the function τ is known except from a parameter, but the average tax rate $\tilde{\tau}$ is observed. In this case, we would need to change the calibration algorithm,

[40] As expected, these equations reduce to the baseline equations when $\tau_i = 0$ for every firm i.

since we now should obtain the unknown parameter of the function τ, $\underline{A}$, and $\overline{A}$ by matching Equations (2.11”), (2.12”), and also (2.F.3) to data.

Regarding these empirical strategies, three final comments are due. First, in Appendix 2.D, we discuss necessary and sufficient conditions for the existence of a solution for the calibration algorithm. We do not make an analogous evaluation here, and thus, a solution for the new calibration problems may not exist, at least for some functions τ. Second, $\mu = \frac{1-\alpha}{LS}$ is still valid here, but now $LS \equiv \frac{Hw}{p\sum_{i=1}^{N}(1-\tau_i)Y_i} = \frac{Hw}{Yp(1-\tilde{\tau})}$ is the labor share of national income net of revenue tax. So, in matching (2.12”) to $\frac{1-\alpha}{LS}$, one can ignore the average tax rate $\tilde{\tau}$ and match $\frac{\overline{A}\Omega}{\underline{B}}$ to $\frac{1-\alpha}{LS_g}$, where $LS_g \equiv \frac{Hw}{Yp}$ is the labor share of national gross income. Third, we use here the baseline Cobb-Douglas production function (2.2). However, in the spirit of Appendix 2.F.1, one can employ an arbitrary well-behaved production function with M factors of production, constant returns to scale, and Hicks-neutral productivity shifter. As before, the only difference would be in the computation of TFP $\overline{A}\Omega$ and average markup μ from data.

2.G Calibrating the capital share parameter for the US using factor income data

In this section, we calibrate α for the US using real-world data. As shown in Appendix 2.A.4, in our model α equals the cost share of capital, that is, $\alpha = \frac{Kr}{Kr+Hw}$.[41] Thus, to calibrate it, we require factor income data, which is not fully available in National Accounts. As pure profit may not be null in our model, capital expenses cannot be calculated as the residual between national and labor income. In other words, we should distinguish capital expenses from pure profit. Barkai (2020) estimates such factor incomes in the US nonfinancial corporate sector for each year between 1984 and 2014. He takes compensation of employees Hw from the National Income and Productivity Accounts (NIPA), but capital expenses Kr are not obtained as a residual. Instead, they are the product of the nominal value

[41] Setting α as the cost share of capital is conceptually equivalent to calibrating the average markup μ to match the pure profit share of national income, defined as $PS \equiv 1 - \frac{Kr+Hw}{Yp}$. Formally, if $\alpha = \frac{Kr}{Kr+Hw}$, $\mu = \frac{1-\alpha}{LS} = \frac{\frac{Hw}{Kr+Hw}}{\frac{Hw}{Yp}} = \frac{1}{\frac{Kr+Hw}{Yp}} = \frac{1}{1-PS} \leftrightarrow PS = \frac{\mu-1}{\mu}$.

of the physical capital stock and a required rate of return, which approximates the leasing cost of one dollar's worth of capital. This required rate is computed in the spirit of Hall and Jorgenson (1967), depending on the cost of borrowing in financial markets, depreciation rates, expected price inflation of capital, and the tax treatment of both capital and debt. Figure 2.G.1 plots the annual estimates for the cost share of capital $\frac{Kr}{Kr+Hw}$. Since the optimal use of factors is probably a better approximation over the long run, we compute its (i) mean, (ii) median, and (iii) steady-state level from an estimated autoregressive process of order one. We find $\alpha = 0.31$ in all three cases, just slightly lower than the standard calibration.

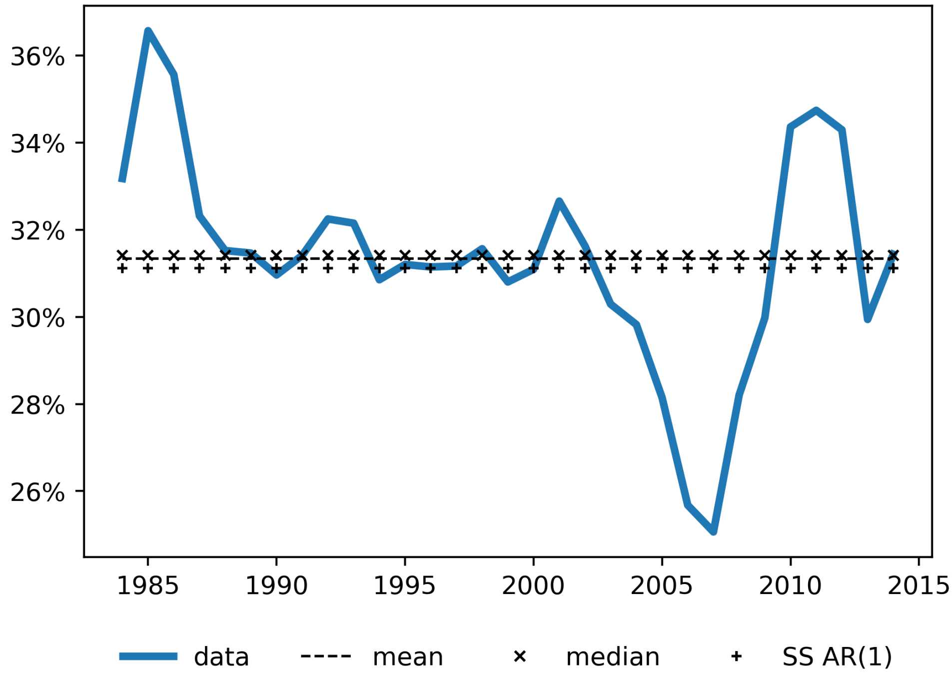


Figure 2.G.1 – Cost share of physical capital in the US nonfinancial corporate sector. *Source*: computations based on data from Barkai (2020).

Using this alternative value for α, we recalculate the average markup $\mu = \frac{1-\alpha}{LS}$. The results are presented in Figure 2.G.2. As expected, the over-time pattern does not change, but these markup estimates are higher than the baseline results of Figure 2.2, by approximately 0.04. For example, in 2012, $\mu = 1.12$ for $\alpha = 1/3$ and $\mu = 1.16$ for $\alpha = 0.31$. Figure 2.G.3 plots the estimates for allocative efficiency Ω based on $\alpha = 0.31$, which are lower and more volatile compared to the baseline results (Figure 2.3). These outcomes align with the observations depicted in Figure 2.1, where a higher markup level points to lower and more volatile Ω.

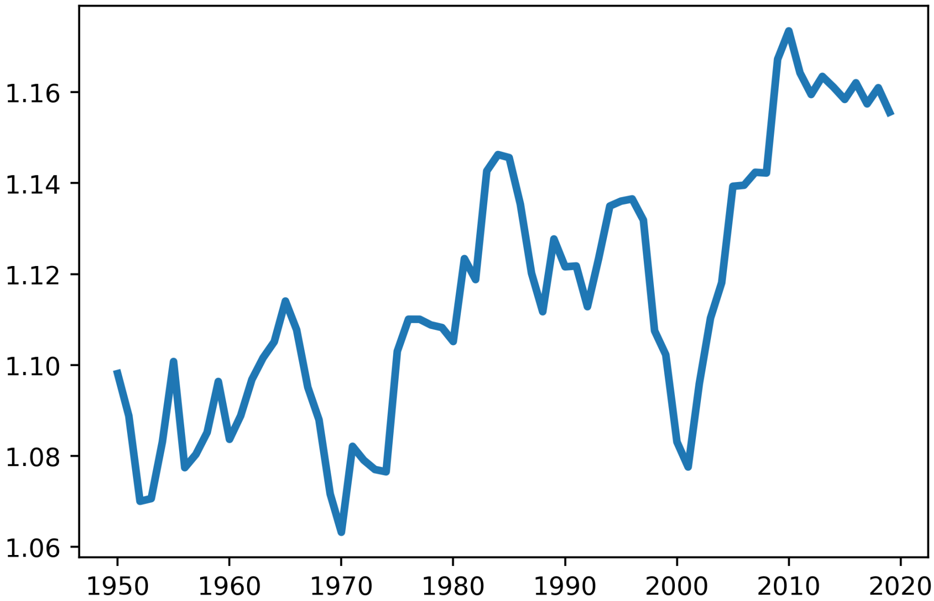


Figure 2.G.2 – Average markup in the United States for $\alpha = 0.31$.

In any case, at least for growth analysis, the differences are not particularly relevant. In Table 2.G.1, we repeat the growth accounting exercise of Table 2.2, but considering $\alpha = 0.31$. The results are practically the same. You can uncover slightly more visible quantitative differences in the TFP decomposition for $k = 5$ (e.g., for 1995-2000 and 2000-2007). However, even in these cases, the fundamental qualitative assessments remain unchanged.

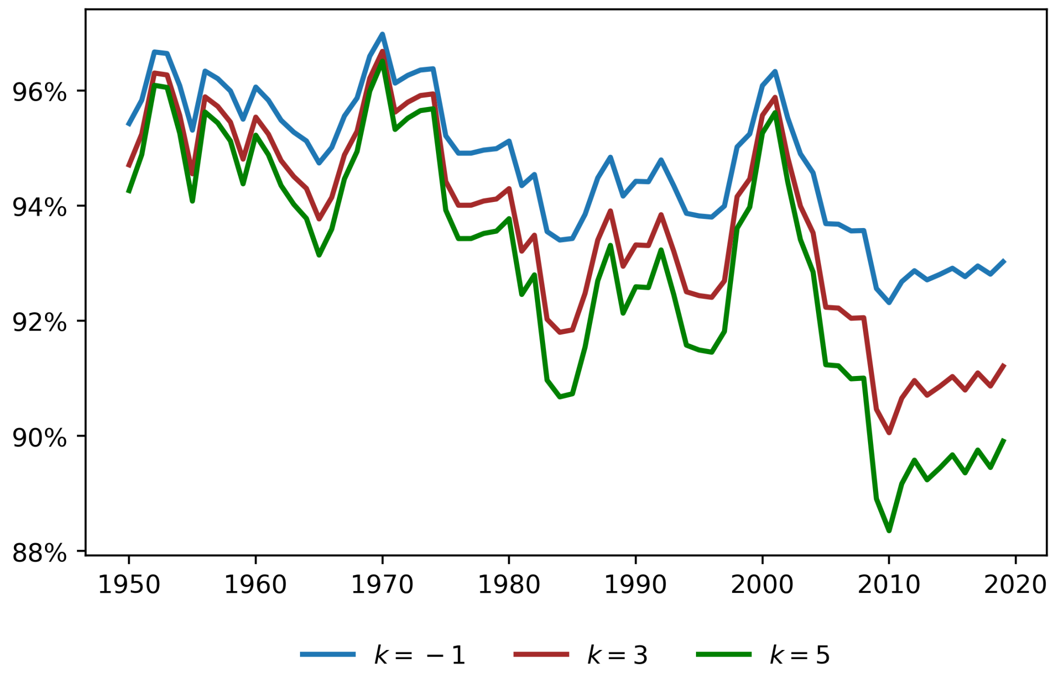


Figure 2.G.3 – Allocative efficiency in the United States for $\alpha = 0.31$.

Table 2.G.1 – Growth accounting for the United States using $\alpha = 0.31$

Period	$\frac{Y}{L}$	Y/L components			Labor-aug. TFP $\left(\overline{A}\Omega\right)^{\beta}$ components					
					$k=-1$		$k=3$		$k=5$	
		$\left(\frac{K}{Y}\right)^{\alpha\beta}$	$\frac{H}{L}$	$\left(\overline{A}\Omega\right)^{\beta}$	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}	$\overline{A}^{\beta}$	Ω^{β}
1954-2019	1.9	0.2	0.5	1.2	1.2	-0.1	1.3	-0.1	1.3	-0.1
1954-2013	1.9	0.2	0.6	1.2	1.3	-0.1	1.3	-0.1	1.3	-0.2
1954-1973	2.6	-0.0	1.0	1.7	1.7	0.0	1.7	0.0	1.7	0.0
1973-1990	1.3	0.3	0.5	0.5	0.7	-0.2	0.8	-0.2	0.8	-0.3
1990-1995	1.6	0.2	0.5	0.9	1.1	-0.2	1.2	-0.3	1.3	-0.3
1995-2000	2.2	0.2	0.3	1.7	1.0	0.7	0.8	1.0	0.6	1.2
2000-2007	2.2	0.3	0.3	1.5	2.1	-0.6	2.3	-0.8	2.5	-0.9
2007-2013	1.3	0.4	0.3	0.6	0.8	-0.2	0.9	-0.4	1.0	-0.5
2013-2019	1.0	-0.1	0.1	1.0	0.9	0.1	0.8	0.1	0.8	0.2

Note: Logarithmic approximation of average annual growth rates (in percent).

2.H Exploring the data sets for level analyses

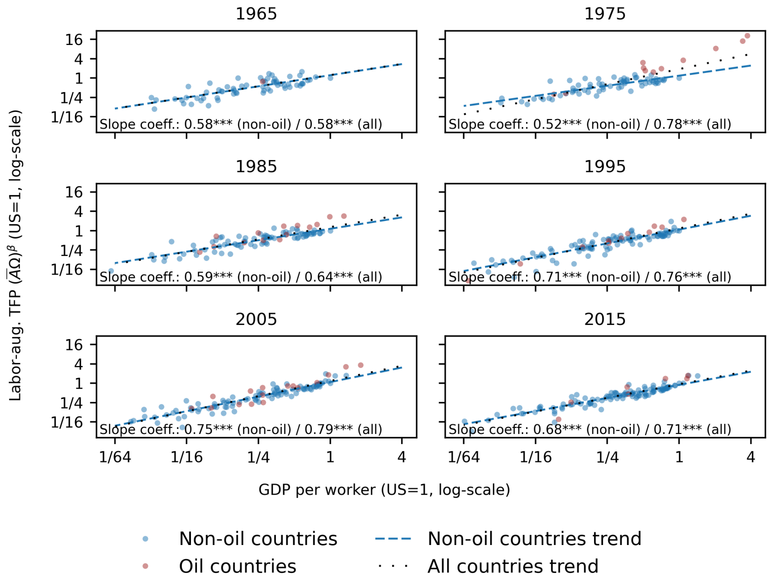


Figure 2.H.1 – Labor-aug. TFP vs. GDP per worker - level data set with greatest coverage, $k = -1$.

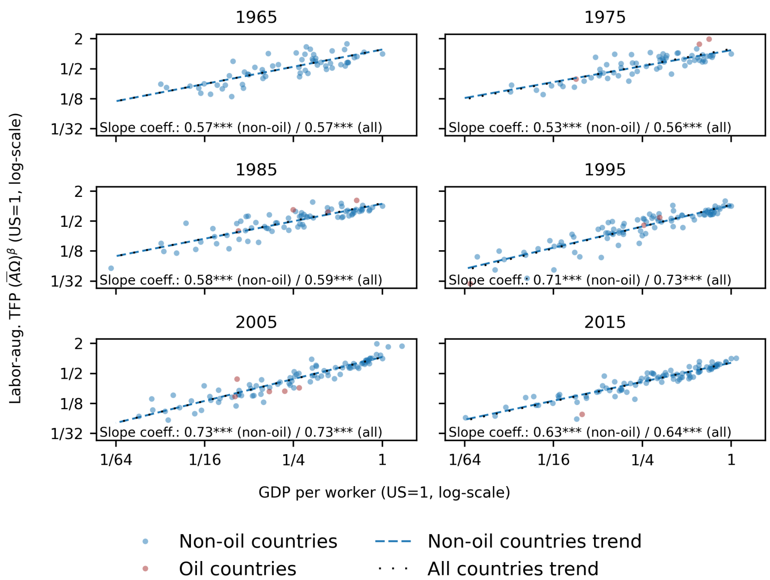


Figure 2.H.2 – Labor-aug. TFP vs. GDP per worker - level data set with greatest coverage, $k = 3$.

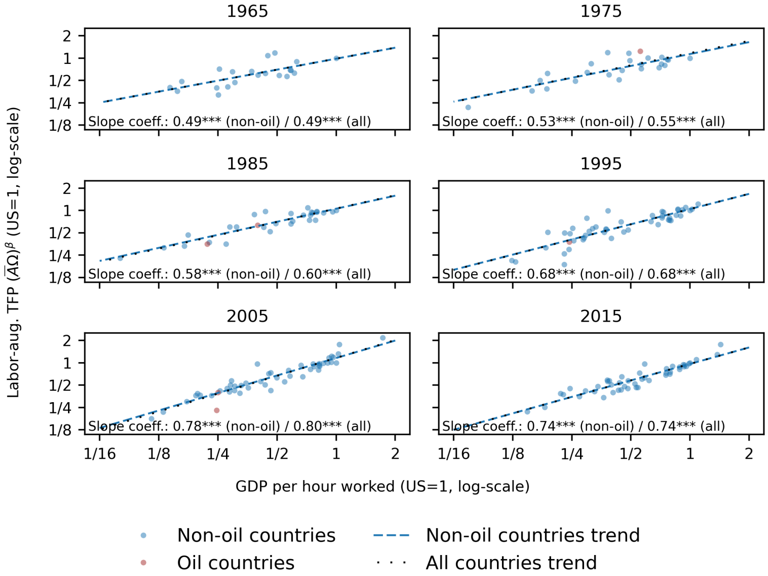


Figure 2.H.3 – Labor-aug. TFP vs. GDP per hour worked - best proxies for level analyses, $k = -1$.

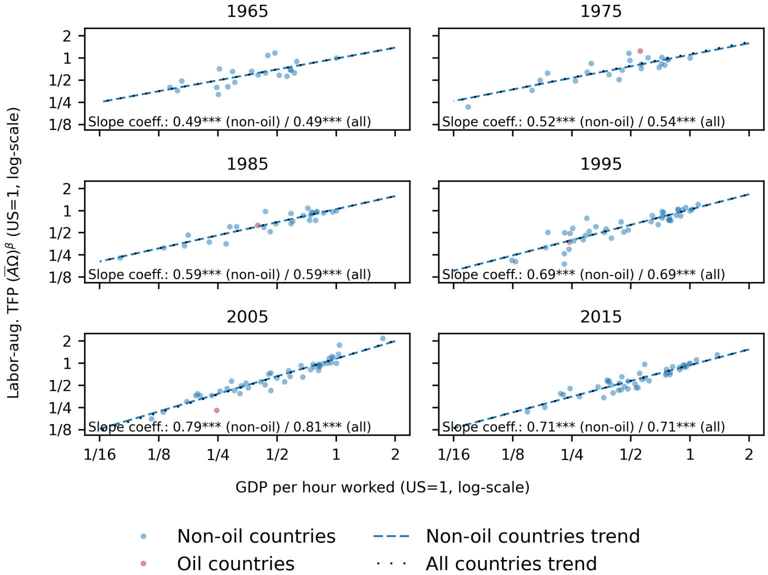


Figure 2.H.4 – Labor-aug. TFP vs. GDP per hour worked - best proxies for level analyses, $k = 3$.

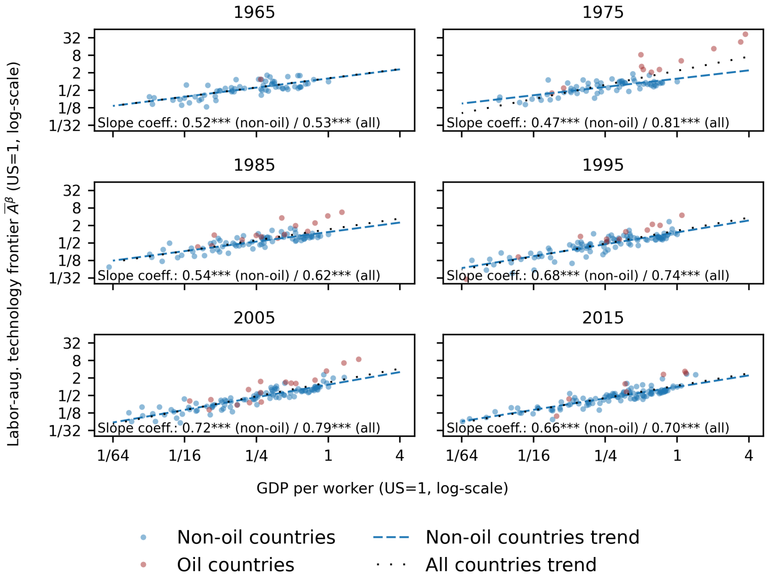


Figure 2.H.5 – Labor-aug. technology frontier vs. GDP per worker - level data set with greatest coverage, $k = -1$.

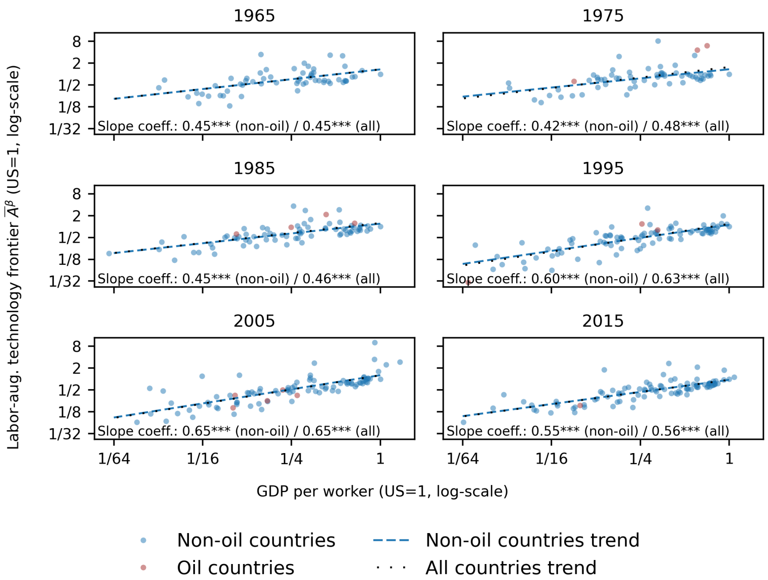


Figure 2.H.6 – Labor-aug. technology frontier vs. GDP per worker - level data set with greatest coverage, $k = 3$.

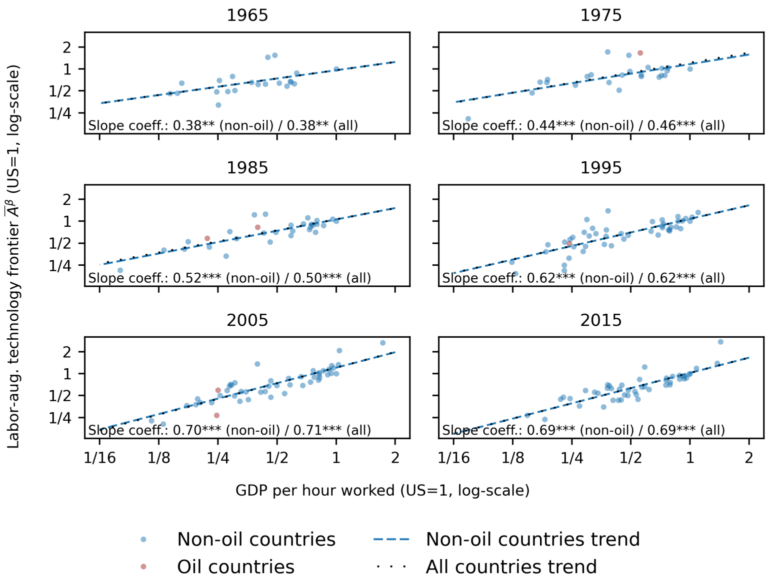


Figure 2.H.7 – Labor-aug. technology frontier vs. GDP per hour worked - best proxies for level analyses, $k = -1$.

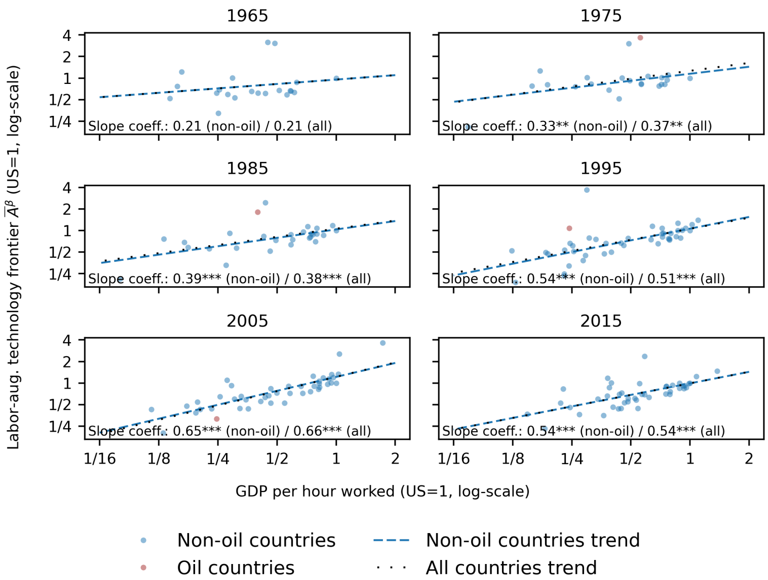


Figure 2.H.8 – Labor-aug. technology frontier vs. GDP per hour worked - best proxies for level analyses, $k = 3$.

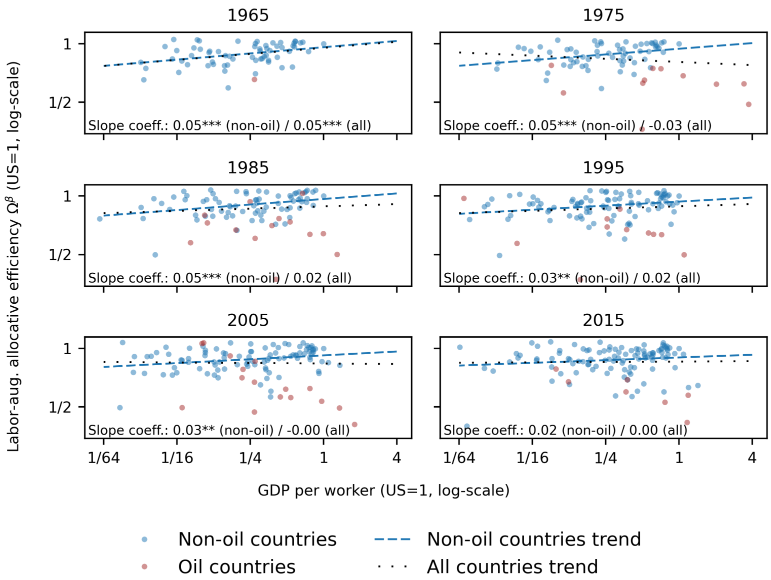


Figure 2.H.9 – Labor-aug. allocative efficiency vs. GDP per worker - level data set with greatest coverage, $k = -1$.

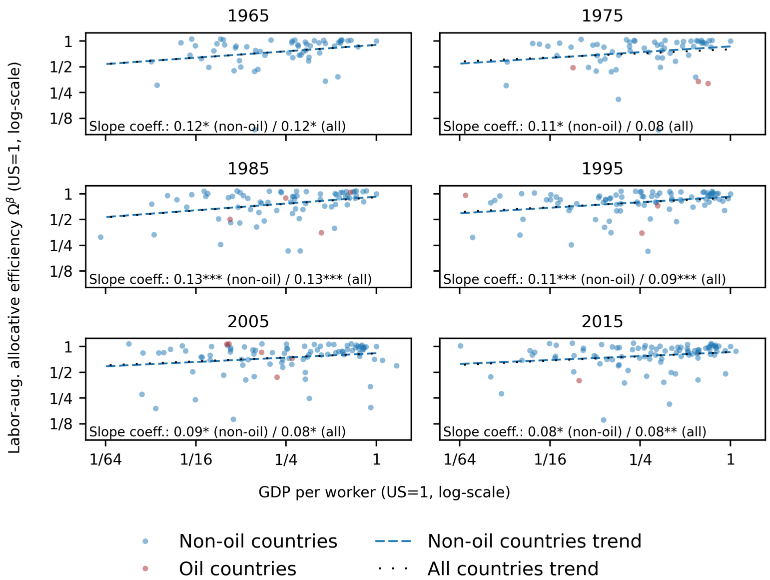


Figure 2.H.10 – Labor-aug. allocative efficiency vs. GDP per worker - level data set with greatest coverage, $k = 3$.

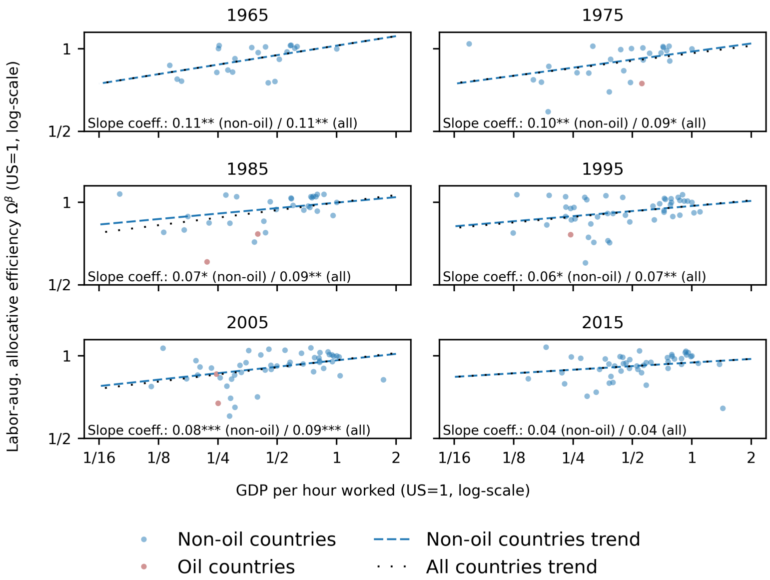


Figure 2.H.11 – Labor-aug. allocative efficiency vs. GDP per hour worked - best proxies for level analyses, $k = -1$.

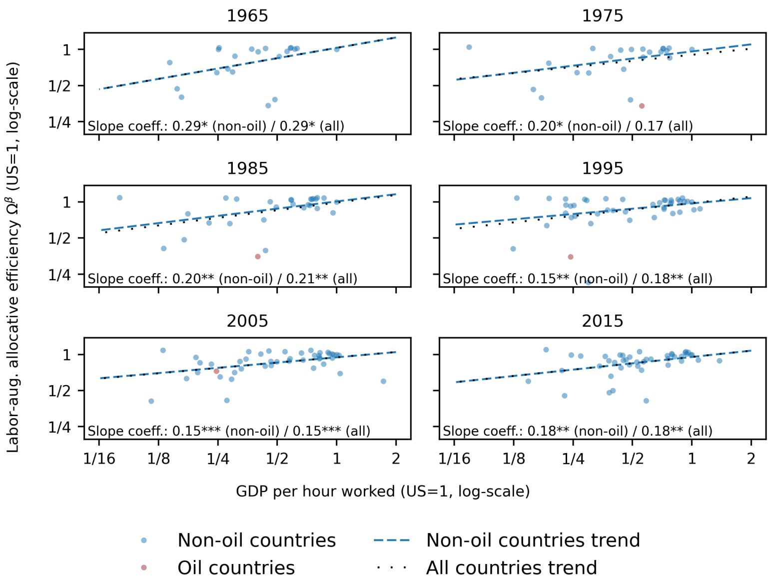


Figure 2.H.12 – Labor-aug. allocative efficiency vs. GDP per hour worked - best proxies for level analyses, $k = 3$.

2.I Exploring the data sets for growth analyses

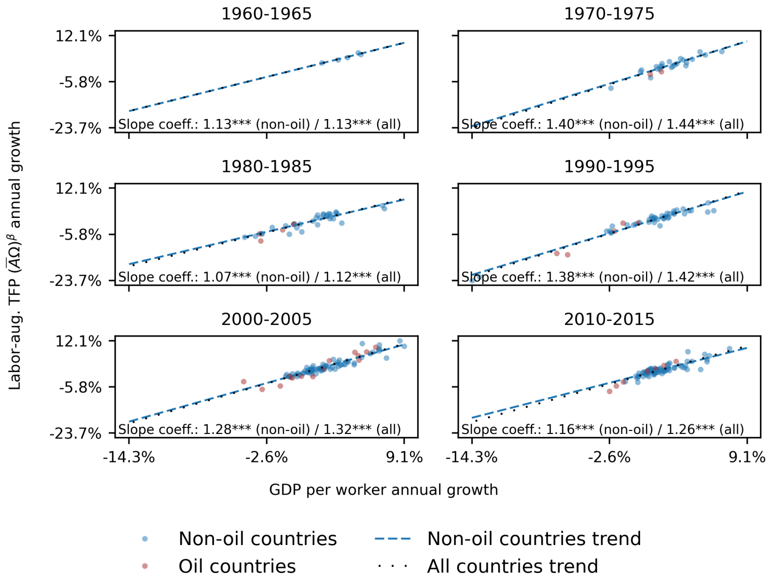


Figure 2.I.1 – Labor-aug. TFP growth vs. GDP per worker growth - growth data set with greatest coverage, $k = -1$.

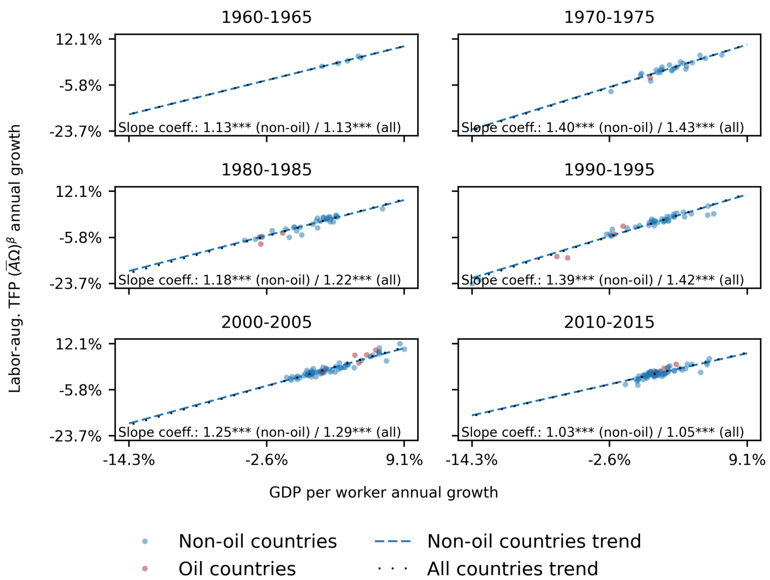


Figure 2.I.2 – Labor-aug. TFP growth vs. GDP per worker growth - growth data set with greatest coverage, $k = 3$.

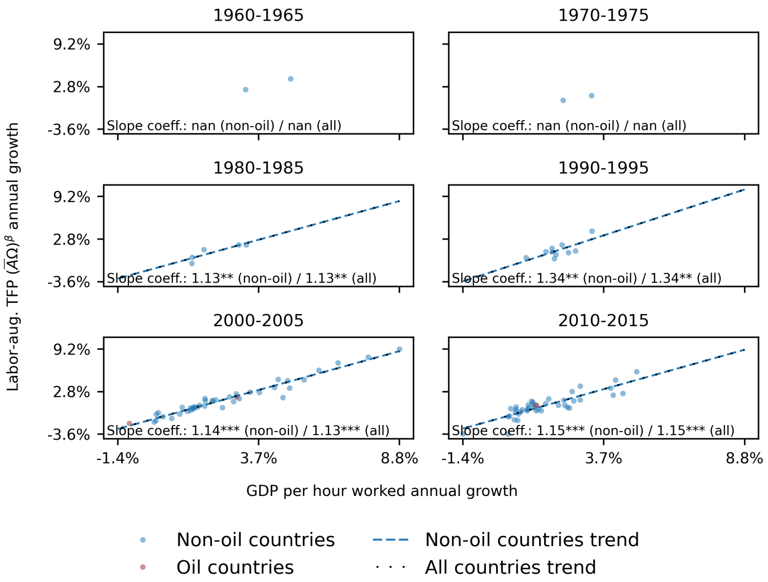


Figure 2.I.3 – Labor-aug. TFP growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = -1$.

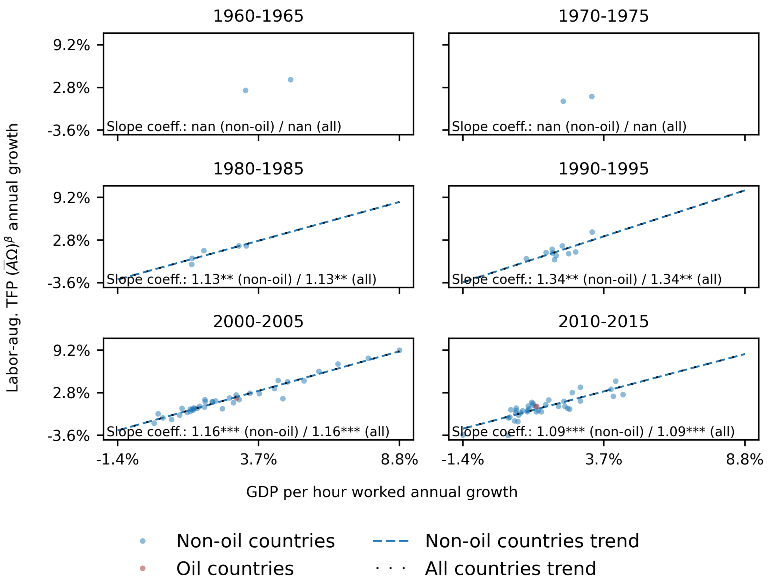


Figure 2.I.4 – Labor-aug. TFP growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = 3$.

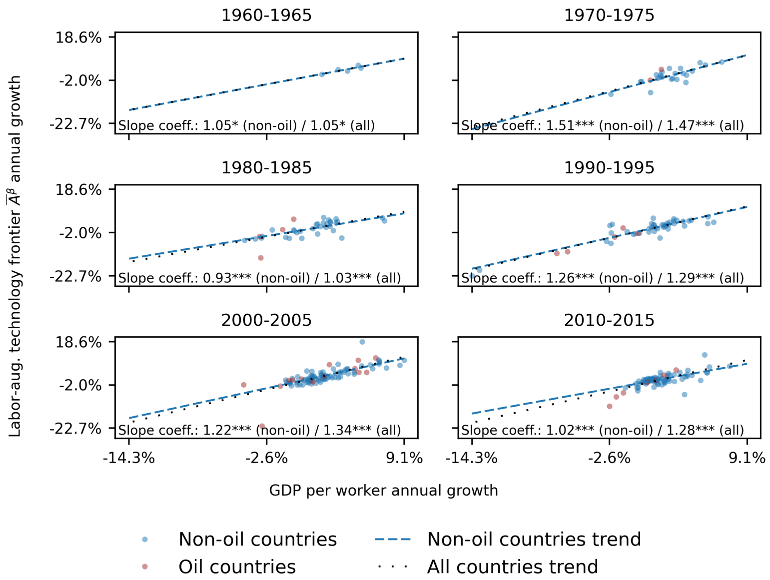


Figure 2.I.5 – Labor-aug. technology frontier growth vs. GDP per worker growth - growth data set with greatest coverage, $k = -1$.

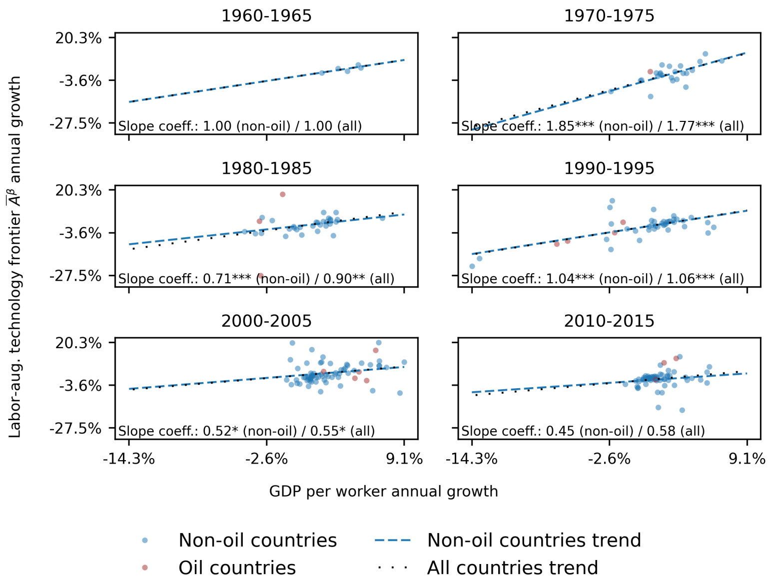


Figure 2.I.6 – Labor-aug. technology frontier growth vs. GDP per worker growth - growth data set with greatest coverage, $k = 3$.

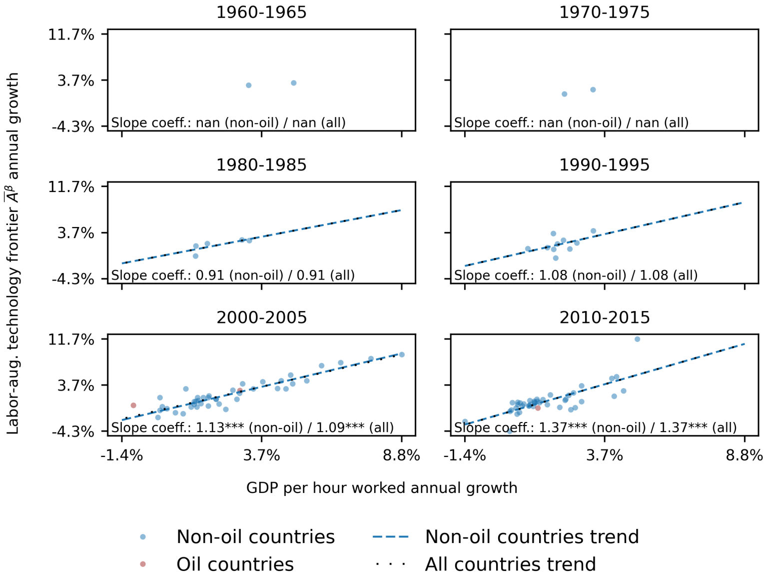


Figure 2.I.7 – Labor-aug. technology frontier growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = -1$.

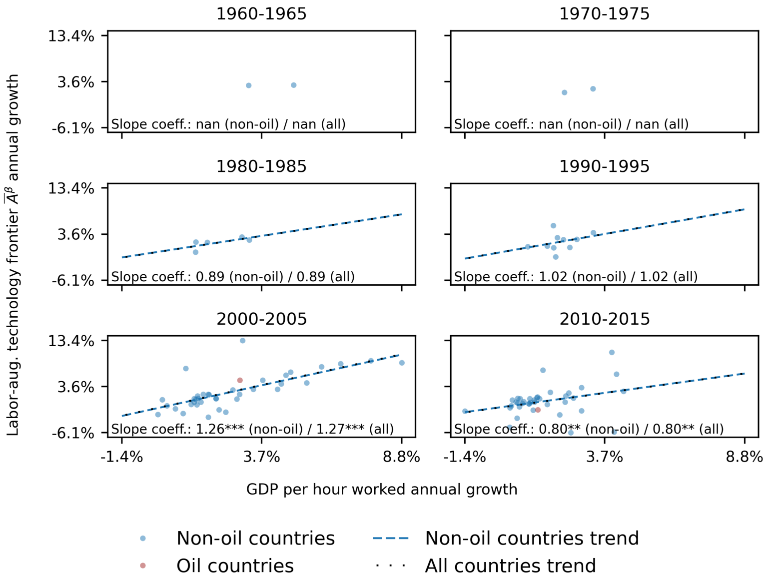


Figure 2.I.8 – Labor-aug. technology frontier growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = 3$.

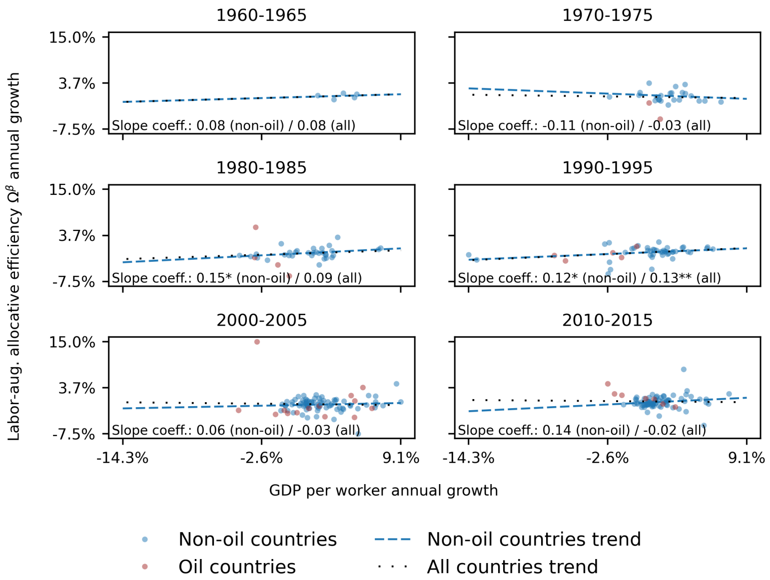


Figure 2.I.9 – Labor-aug. allocative efficiency growth vs. GDP per worker growth - growth data set with greatest coverage, $k = -1$.

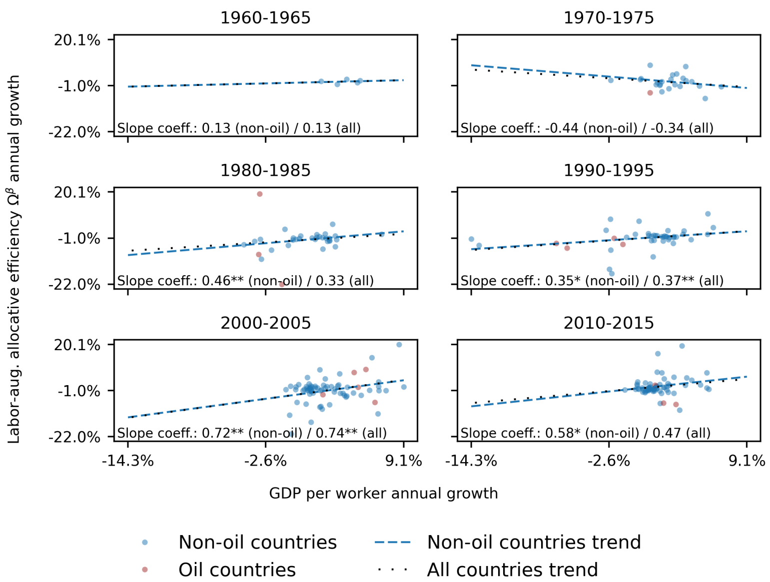


Figure 2.I.10 – Labor-aug. allocative efficiency growth vs. GDP per worker growth - growth data set with greatest coverage, $k = 3$.

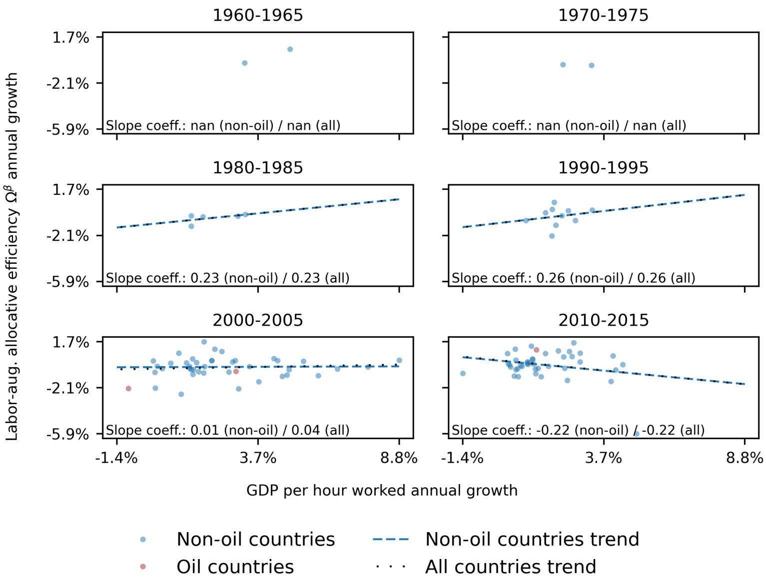


Figure 2.I.11 – Labor-aug. allocative efficiency growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = -1$.

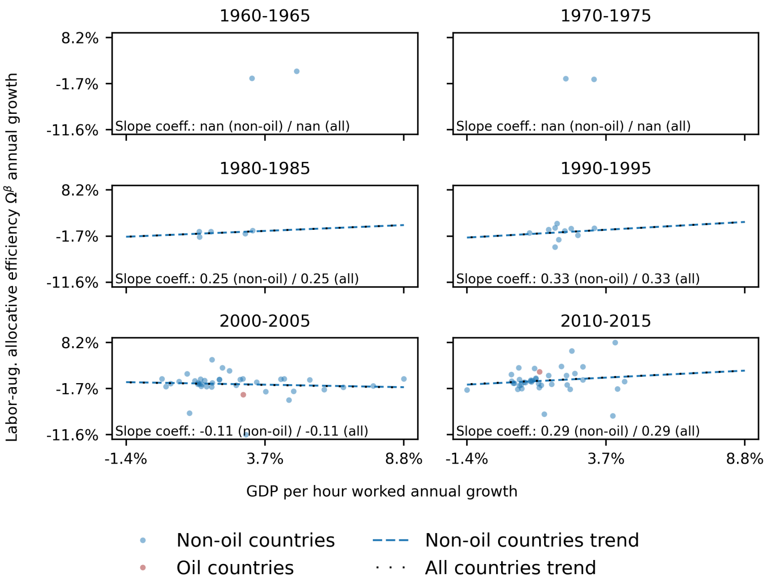


Figure 2.I.12 – Labor-aug. allocative efficiency growth vs. GDP per hour worked growth - best proxies for growth analyses, $k = 3$.

2.J Measures of success for different samples

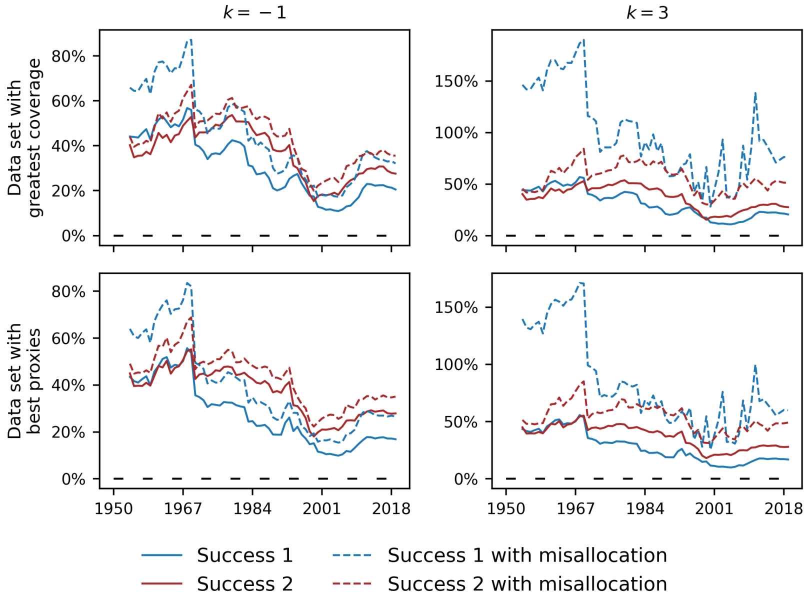


Figure 2.J.1 – Measures of success, same non-oil countries' sample across plots in each given year.

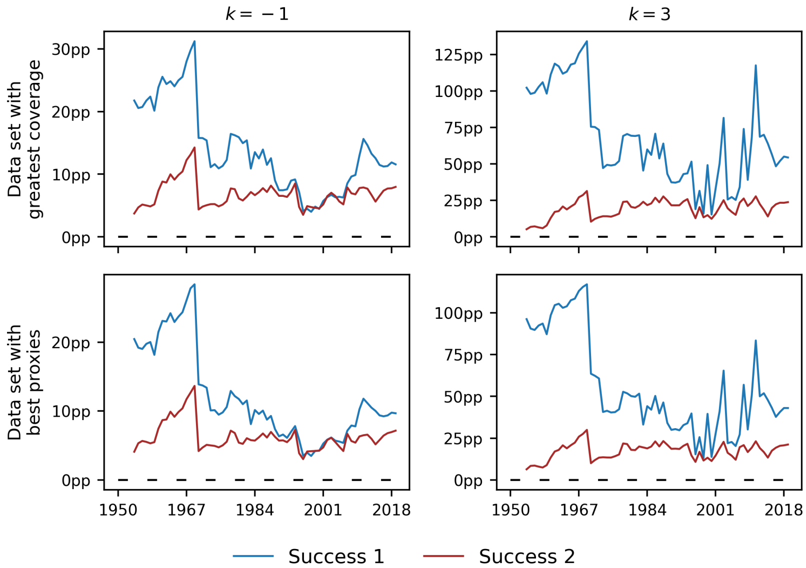


Figure 2.J.2 – Success gains due to misallocation, same non-oil countries' sample across plots in each given year.

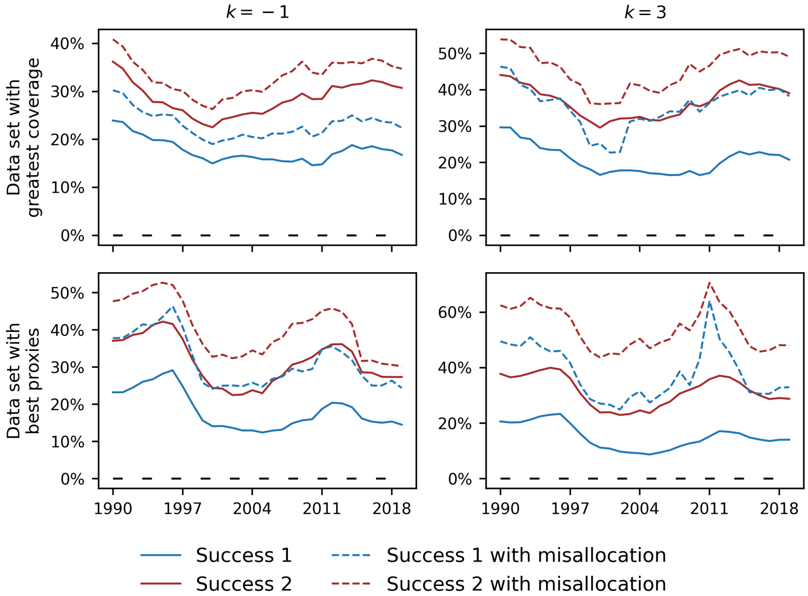


Figure 2.J.3 – Measures of success, time-invariant non-oil countries' sample.

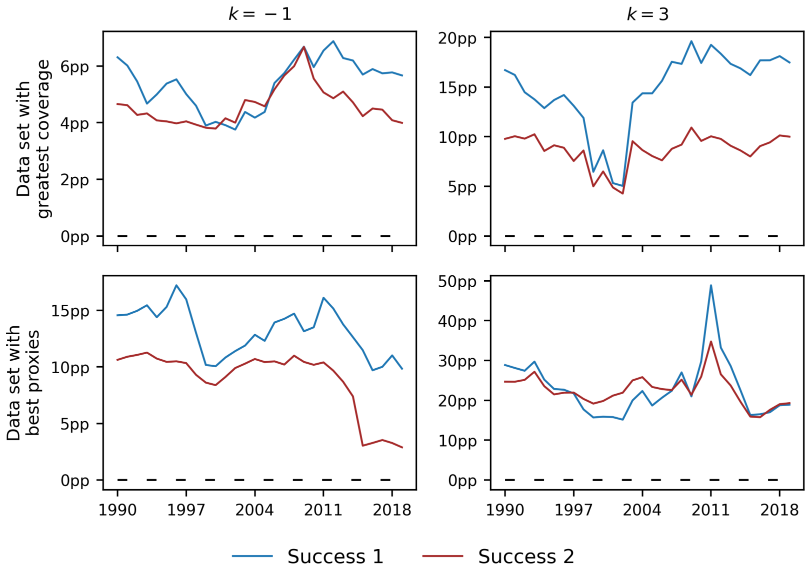


Figure 2.J.4 – Success gains due to misallocation, time-invariant non-oil countries' sample.

3 Disentangling Brazilian TFP: the role of misallocation in recent economic cycles

3.1 Introduction

Aggregate total factor productivity (TFP) is crucial in explaining economic performance both across countries and over time (Klenow; Rodríguez-Clare, 1997; Caselli, 2005; Caselli, 2016; Jones, 2016; Bergeaud; Cette; Lecat, 2018; Crafts; Woltjer, 2021). However, TFP is usually a residual, "a measure of our ignorance," capturing the unexplained portion of such economic outcomes. As an aggregate measure, a natural approach to disentangle the TFP growth is to decompose it into increases in firms' productivity (technology component) and composition effects (allocative efficiency component). This decomposition of the aggregate productivity is not unique, as one can find different methods in the growth-accounting literature (Basu; Fernald, 2002; Petrin; Levinsohn, 2012; Baqaee; Farhi, 2020). Besides their differences, all these methods rely on highly general models that do not impose any specific market structure but require microdata that are hardly available for all firms in the economy. The standard solution is to use (i) sectoral data that span the economy (Basu; Fernald, 2002), (ii) firm-level data that do not span the economy (Petrin; Levinsohn, 2012), or (iii) both (Baqaee; Farhi, 2020). We follow an alternative, complementary, approach that requires minimal microdata by relying on stronger model assumptions.

More precisely, we employ the static Cournot model of Chapter 2, where firms have distinct productivity and consequently charge different markups, leading to inefficient allocation of resources as the marginal products would not be equalized across firms. Interestingly, this model primarily relies on macroeconomic data for calibration, which is only possible because of its stronger assumptions. On the one hand, the model exclusively assesses market-power-driven misallocation. On the other hand, the model is less flexible than other related oligopoly models such as Edmond, Midrigan and Xu (2015) and Loecker, Eeckhout and Mongey (2021), in three important aspects. First, instead of having firms producing distinct goods in each sector over a continuum of sectors, we suppose there is only one sector where firms produce the same good. Second, there are no fixed costs. Third, we assume free entry among low-productivity firms, supposing the number of inefficient firms is sufficiently large to the extent that some will not be active.

As a result of these stronger assumptions, allocative efficiency and other key model expressions do not depend on parameters such as the price elasticity of demand and the number of potential or active firms. They require only the empirical distribution of firms' productivity, which we assume is truncated Pareto. In Chapter 2, we used the same density, testing different values for the Pareto shape parameter and searching the distributional support by matching the aggregate TFP and the cost-weighted average of firm-level markups. We follow a similar strategy here, but we also estimate the shape parameter, using the number of active firms multiplied by the Herfindahl–Hirschman concentration index (HHI) of the labor market as our third target moment.

To compute the target moments using real-world data, we first parameterized the firms' production function, which is Cobb-Douglas with constant returns to scale, utilizing both capital and labor as inputs. The capital share parameter α is calibrated at 0.39 based on cost share data. Labor expenses are readily obtained from National Accounts, while capital expenses are computed as outlined in Barkai (2020), estimating a required rate of return in the spirit of Hall and Jorgenson (1967). With α determined, the first two moments are easily obtained from standard macroeconomic data. On the one hand, the cost-weighted average markup equals $1 - \alpha$ divided by the labor share of national income, which is consistent with Hall (1988) and the recent literature that uses his results to gauge firm-level markups (Loecker; Warzynski, 2012; Loecker; Eeckhout, 2018; Loecker; Eeckhout; Unger, 2020; Traina, 2018; Calligaris; Criscuolo; Marcolin, 2018; Autor et al., 2020). On the other hand, the TFP is backed out as a residual in the aggregate production function, given labor and capital stock data, both adjusted by varying utilization. Rather than merely multiplying the capital stock by a capacity utilization indicator, we consider a "Cobb-Douglas capacity utilization share parameter." The underlying assumption is that capital and labor utilization is simply a rescaling of an observable measure of intensity in input usage, in line with the methodology of Basu, Fernald and Kimball (2006) and Basu et al. (2013). We estimate this parameter by employing, for a time series, the simplified model of the dynamic panel literature on production function estimation (Blundell; Bond, 2000) presented in Ackerberg, Caves and Frazer (2015).

However, obtaining a representative third moment for the entire economy proves challenging even with comprehensive firm-level data, as it requires pursuing the difficult task of defining the relevant market for each firm in the economy (Berry; Gaynor; Morton, 2019; Syverson, 2019; Benkard; Yurukoglu; Zhang, 2021). Given that scenario, we seek a weaker moment match in this case, trying only to approximate the shape of its time series by focusing on the normalized moment. By doing that, we do not need to estimate it precisely as both its mean and variance are irrelevant, allowing us to employ relatively aggregate employment data and simple assumptions regarding the relevant markets (e.g., there is only one national market) to compute it.

Employing this empirical strategy, we decompose aggregate TFP. We estimate the allocative efficiency, measured by the distance from optimal allocation, using solely the average markup and the concentration measure of the labor market. TFP data is used only to pin down the technology component of TFP. Hence, as in Chapter 2, the residual of the production function is not the TFP itself but rather only its technology component, which is a cleaner residual as it is free of misallocation effects. Interestingly, as demonstrated in Chapter 2, allocative efficiency is a strictly decreasing function of the average markup and, thus, strictly increases with the labor share of national income. In particular, the allocation becomes optimal when the markup converges to one. This is intuitive as an average markup closer to the competitive unitary level suggests a less distorted economy, indicating closer proximity to optimal allocation.

We apply this methodology to Brazil between 2000 and 2019. Overall, allocative efficiency improved, reflecting the observed increase in the labor income share and, thus, the estimated decrease in the average markup. This is a sharp contrast with most developed countries, which over the last decades experienced decreasing labor share and increasing average markup (Calligaris; Criscuolo; Marcolin, 2018;

Loecker; Eeckhout, 2018; Autor et al., 2020). Accordingly, studying the US, Baqaee and Farhi (2020) find that allocative efficiency, as measured by the distance from optimal allocation, deteriorated in 2015 relative to 1997. Additionally, we find that the cycles in Brazilian TFP are mainly due to allocative efficiency, with the economic boom in the mid-2000s being primarily attributed to efficiency gains. The technology component of TFP grows much more steadily, around 0.8-0.9% per year, suggesting it reflects the structural characteristics of the economy. Since allocative efficiency could not increase or decrease indefinitely, this annual technology growth of 0.8-0.9% can be seen as the current structural, long-run, growth level of Brazilian TFP.

We argued in Chapter 2 that a lower Pareto shape parameter yields more conservative results compared to standard growth-accounting exercises, as the estimated allocative efficiency becomes less volatile. Thus, in a robustness exercise, instead of calibrating it, we follow the baseline approach of Chapter 2, setting the shape parameter to the lowest level that remains consistent with the reasonable and usual assumption that high-productivity firms are relatively scarce. Given this shape parameter, we estimate the model again, using *only* macroeconomic data as the third moment is not required in this case, making it robust to possible measurement errors in labor market concentration data.[1] We find qualitatively similar results.

In another robustness exercise, we set the capital share parameter α to the standard value of 1/3 or 0.41 instead of our baseline calibration ($\alpha = 0.39$). Additionally, we experiment with a higher labor income share by allocating all self-employment income (mixed income) to labor. In the baseline case, we allocate mixed income to labor and non-labor in the same proportions as the rest of the economy, following the preferred method of the Penn World Table 10.01 (Feenstra; Inklaar; Timmer, 2015). In both scenarios, we again find qualitatively similar

[1] Our primary quantification strategy just requires sound estimates of the overall trends in labor market concentration. However, even this weaker requirement may not be fulfilled. For instance, Benkard, Yurukoglu and Zhang (2021) compute product market concentration in the US between 1994 and 2019 for (i) "[...] narrowly defined product markets as would be defined in an antitrust setting" (Benkard; Yurukoglu; Zhang, 2021, p.1) and (ii) broader sector levels (closer to what we do), finding concentration is decreasing in the first case and increasing in the second.

results.

In a final robustness exercise, we propose a different strategy, making an assumption about the distribution of firms' market share instead of choosing a productivity density. Consistent with the empirical evidence supporting Zipf's law for firm size (Okuyama; Takayasu; Takayasu, 1999; Axtell, 2001; Fujiwara et al., 2004; Luttmer, 2007; Gabaix; Landier, 2008; Giovanni; Levchenko; Ranciere, 2011; Giovanni; Levchenko, 2013; Silva et al., 2018), we assume the market share distribution is truncated Lomax, which is a particular case of the truncated Pareto distribution type II, setting its shape parameter close to one. As before, we search the remaining parameters, including the Lomax scale parameter, by matching the model predictions to the same three data moments used in the baseline case. We find practically identical results.

Related literature. We employ the static Cournot model of Chapter 2, which relates to several papers that embed oligopoly market structures in macroeconomic models (Bernard et al., 2003; Atkeson; Burstein, 2008; Edmond; Midrigan; Xu, 2015; Peters, 2020; Loecker; Eeckhout; Mongey, 2021; Wang; Werning, 2022; Edmond; Midrigan; Xu, 2022). That model is notable for its requirement of minimal microdata for calibration, achieved through reliance on stronger assumptions. This chapter contributes to the previous chapter's work in several dimensions. First, we derive the model's expressions for (i) the HHI of the labor market and (ii) the HHI of the product market. Second, using that expression for the labor market concentration, we estimate the Pareto shape parameter from the data. In contrast, in Chapter 2, we explored some scenarios by assigning different values to the shape parameter. Third, in a robustness exercise, we extend the analysis beyond a Pareto productivity distribution and assume that firms' market share follows a Lomax distribution. In Chapter 2, we established necessary and sufficient conditions for achieving an exact match of the first two target moments, namely, aggregate TFP and average markup. We first derived general results by examining an arbitrary truncated distribution, which we subsequently applied to delineate the specific conditions for the Pareto case. We adopt a similar strategy here, utilizing those general results to derive analogous conditions for the Lomax case. Finally, we account for the potential underutilization of inputs and adjust capital and labor using a capacity utilization index, in the spirit of Basu, Fernald and Kimball (2006) and Basu et al. (2013).

Given our goal of disentangling TFP growth, this paper is closely related to the growth-accounting literature that decomposes it into technology and allocative efficiency components (Basu; Fernald, 2002; Petrin; Levinsohn, 2012; Baqaee; Farhi, 2020). Nevertheless, this literature usually relies on highly flexible models, whose quantification requires extensive microdata that are hardly available. Additionally, this literature typically adopts a different concept of allocative efficiency. As pointed out by Baqaee and Farhi (2020, p.107), "[...] the growth-accounting notion of changes in allocative efficiency due to the reallocation of resources to more or less distorted parts of the economy over time is very different from the misallocation literature's notion of allocative efficiency measured as the distance to the Pareto-efficient frontier." Conversely, we employ the model of Chapter 2, which relies on stronger assumptions and, consequently, requires mainly macroeconomic data to gauge allocative efficiency, as measured by the distance from optimal allocation.

Finally, our work shares common ground with the literature on misallocation (Restuccia; Rogerson, 2008; Hsieh; Klenow, 2009; Restuccia; Rogerson, 2013).[2] However, in this literature, misallocation is a result of exogenous wedges, whose estimation requires extensive firm-level data, similar to growth-accounting methods. In contrast, misallocation is endogenous in the model of Chapter 2, emerging as an equilibrium outcome. Furthermore, in this literature, the typical goal is to gauge the importance of misallocation in explaining cross-country TFP differences, usually within the manufacturing sector due to data availability, while we focus on economy-wide TFP evolution over time in a given country.

The remainder of the paper proceeds as follows. Section 3.2 briefly reviews the model of Chapter 2. Section 3.3 presents our quantification strategy, which requires data and parameters obtained for Brazil in Section 3.4. Section 3.5 discusses the baseline empirical results for Brazil, while Section 3.6 conducts some robustness checks. Finally, Section 3.7 concludes.

[2] For papers on misallocation in Brazil, refer to Busso, Madrigal and Pagés (2013) and Vasconcelos (2017) for the manufacturing industries and Vries (2014) for the retail sector.

3.2 Model

We use the static Cournot model of Chapter 2, where firms have different productivity levels and consequently charge distinct markups, generating misallocation of resources. Formally, in a closed economy, N potential entrant firms produce a single good. The price elasticity of demand for this good is strictly negative, with its absolute value denoted by η, where $1 < \eta < \infty$. The output of firm $i \in \{1, 2, ..., N\}$ is given by the Cobb-Douglas function

$$Y_i = A_i K_i^{\alpha} L_i^{1-\alpha} \tag{3.1}$$

where $K_i \geq 0$ is the capital stock, $L_i \geq 0$ is the labor employed, and $A_i > 0$ is a productivity parameter, all for firm i, while $\alpha \in (0, 1)$. Let $\underline{\underline{A}} \equiv \min_i\{A_i\}$ and $\overline{A} \equiv \max_i\{A_i\}$ be the technology frontier of this economy, with $0 < \underline{\underline{A}} < \overline{A} < +\infty$.

These firms compete a la Cournot, all taking the wage $w > 0$ and the rental cost of capital $r > 0$ as given. There are no fixed costs, but some firms may not be active due to the assumption of homogeneous goods. As a consequence, we need to employ an entry stage to obtain the set of active firms in equilibrium, when (i) each active firm has non-negative profits and (ii) non-active firms would make strictly negative profits if they entered the market. However, this equilibrium is usually not unique (Atkeson; Burstein, 2008; Edmond; Midrigan; Xu, 2015; Loecker; Eeckhout; Mongey, 2021). To avoid multiple equilibria, we discard any equilibrium where a non-active firm has a lower marginal cost than an active firm. As a consequence, in the (unique) refined equilibrium, there exists a firm with productivity $\underline{A}$ serving as the cutoff for active firms, such that firm i is active if and only if $A_i \geq \underline{A}$. Finally, we consider free entry among low-productivity firms, supposing the number of such firms is sufficiently large to the extent that some should be inactive. This implies the profit of the firm with productivity $\underline{A}$ should be, in some sense, low. We assume it is approximately null.

Given that setup, from Chapter 2,

$$s(A_i) \approx \eta \left(1 - \underline{A}/A_i\right) \tag{3.2}$$

$$\eta \approx \frac{1}{N_a \left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]} \tag{3.3}$$

$$Y = \overline{A}\Omega K^{\alpha}L^{1-\alpha} \tag{3.4}$$

$$\Omega \approx \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1-\underline{A}/A)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]} \tag{3.5}$$

$$\mu \approx \frac{\overline{A}\Omega}{\underline{A}} \tag{3.6}$$

where $s(A_i)$ is the market share of a firm with productivity A_i, $A_i \geq \underline{A}$, which is clearly a strictly increasing function of A_i. $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|A \geq \underline{A}\right) = \sum_{A\geq\underline{A}} h(A)\frac{g(A)}{1-G(\underline{A})}$ is the expected value of a function h over active firms under the empirical distribution and $N_a \equiv N\left(1-G(\underline{A})\right)$ is the number of active firms, with $g(A)$ being the empirical probability of A and $G(A) = \sum_{a<A} g(a)$ the empirical cumulative distribution function. Moreover, $Y \equiv \sum_{i=1}^{N} Y_i$ is the aggregate output, $K \equiv \sum_{i=1}^{N} K_i$ is the aggregate capital stock, $L \equiv \sum_{i=1}^{N} L_i$ is the aggregate amount of labor employed, and $\overline{A}\Omega$ is the aggregate TFP. Note $\Omega \in (0,1]$, which is intuitive as Ω measures allocative efficiency by the distance of aggregate TFP $\overline{A}\Omega$ from its optimal level, $\overline{A}$. After all, as firms produce homogeneous goods, the optimal allocation entails assigning all inputs to the most productive firm, when $\Omega = 1$ and $Y = \overline{A}K^{\alpha}L^{1-\alpha}$. Finally, being μ_i the markup of firm i, $\mu \equiv \sum_{i=1}^{N}\left(\frac{L_i w + K_i r}{Lw+Kr}\right)\mu_i$ is the cost-weighted average of firm-level markups.

Interestingly, we showed in Chapter 2 that Ω improves when $\underline{A}$ increases due to the exit of less productive active firms from the market, suggesting that allocative efficiency is closely related to productivity dispersion among active firms. Moreover, since any allocation of resources is optimal when there is no productivity dispersion, $\Omega \to 1$ as $\underline{A} \to \overline{A}$, which corresponds to a market in perfect competition, where all active firms have unitary markups and null profits, with $N_a \to \infty$.

Relying on the previous chapter's results, we derive expressions for two measures of concentration: (i) the Herfindahl–Hirschman index (HHI) of the labor market $HHI_L \equiv \sum_{i=1}^{N}(L_i/L)^2$ and (ii) the HHI of the product market $HHI_s \equiv \sum_{i=1}^{N} s(A_i)^2$. As shown in Appendix 3.A,

$$N_a HHI_L \approx \frac{\mathrm{E}_a\left[(\underline{A}/A)^2\left(1-\underline{A}/A\right)^2\right]}{\left\{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]\right\}^2} \tag{3.7}$$

$$N_a HHI_s \approx \frac{\mathrm{E}_a\left[\left(1-\underline{A}/A\right)^2\right]}{\left[\mathrm{E}_a\left(1-\underline{A}/A\right)\right]^2} \approx \frac{1-1/\mu}{1-\mathrm{E}_a\left(\underline{A}/A\right)} \tag{3.8}$$

In Chapter 2, we also presented a version of the model with a continuum of firms of mass N. In such circumstances, the standard practice would be to assume these null-measure firms ignore the impacts of their decisions on aggregate outcomes even though they exist (e.g., macroeconomic models of monopolistic competition). Formally, firms would consider $\partial Y/\partial Y_i = 0$ even though $\partial Y/\partial Y_i > 0$, with $\partial Y/\partial Y_i = 1$ for homogeneous goods. However, we followed a different approach, supposing firms pay at least some attention to them, with $\partial Y/\partial Y_i = q \in (0, 1]$ for each firm $i \in [0, N]$. In this case, all the above equations, including (3.7) and (3.8), would hold *exactly* if (i) η is replaced by η/q and (ii) integrals are used instead of sums (e.g., $Y \equiv \int_0^N Y_i di$ and $\mathrm{E}_a(h(A)) = \int_{\underline{A}}^{\overline{A}} h(A) \frac{g(A)}{1-G(\underline{A})} dA$, where g is now a density function).[3] They hold exactly because now the profit of the least productive active firm is precisely zero under the free-entry assumption, since marginal adjustments in the number of active firms are allowed in this case. Since the two models have essentially the same equations, they have similar properties: (i) $\Omega \in (0, 1]$, (ii) Ω strictly increases with $\underline{A}$, and (iii) $\Omega \to 1$ as $\underline{A} \to \overline{A}$.

3.3 Quantification strategy

In this section, we outline the calibration strategy. Our primary empirical objective is to compute allocative efficiency Ω, which requires solely the distribution of firms' productivity (Equation (3.5)). Thus, we aim to utilize real-world data to pin down the distributional parameters. We propose a two-stage procedure. In the first stage, we employ the calibration algorithm of Chapter 2, finding, for a given distributional shape, $\underline{A}$ and $\overline{A}$ by matching aggregate TFP $\overline{A}\Omega$ and average markup μ to data. In the second stage, relying on this first-stage algorithm, we recover the distributional shape from the labor concentration measure $N_a HHI_L$. We present this two-stage calibration procedure in the following. However, first, we discuss our distributional assumption and the computation of those moments using real-world data.

[3] In the discrete model, $\eta > 1$ is sufficient for the Second-Order Condition (SOC) for firms' profit maximization to hold. In the continuous case, $\eta > 0$ or $q \in (0, 1]$ should be low.

3.3.1 Distributional assumption

As in Chapter 2, we assume firm productivity is truncated Pareto distributed with shape parameter $k \neq 0$, consistent with the continuous version of the model. Under this distribution, we demonstrated in the previous chapter that

$$\mathrm{E}_a\left((\underline{A}/A)^j\right) = \begin{cases} \left(\frac{k}{k+j}\right)\left(\frac{\tilde{A}^{k+j}-1}{\tilde{A}^{k+j}-\tilde{A}^j}\right) & \text{, if } k+j \neq 0 \\ \left(\frac{k\tilde{A}^k}{\tilde{A}^k-1}\right)\ln\tilde{A} & \text{, if } k+j = 0 \end{cases} \tag{3.9}$$

for $k \neq 0$, $j \in \mathbb{N} \setminus \{0\}$, and $\tilde{A} \equiv \overline{A}/\underline{A} > 1$.

3.3.2 Computing the target moments using real-world data

We consider three target moments: (i) the aggregate TFP $\overline{A}\Omega$, (ii) the cost-weighted average of firm-level markups μ, and (iii) the labor concentration measure $N_a HHI_L$. As discussed in Chapter 2, the first two moments can be easily obtained using standard macroeconomic data and the parameter α. On the one hand, TFP is backed out as a residual in the aggregate production function (3.4): $\overline{A}\Omega = \frac{Y}{K^\alpha L^{1-\alpha}}$. On the other hand, $\mu = \frac{1-\alpha}{LS}$, with LS being the labor share of national income. However, obtaining a representative $N_a HHI_L$ for the entire economy proves challenging even with comprehensive firm-level data, as it requires pursuing the difficult task of defining the relevant market for each firm in the economy (Berry; Gaynor; Morton, 2019; Syverson, 2019; Benkard; Yurukoglu; Zhang, 2021). This is why we will pursue a weaker moment match for $N_a HHI_L$, trying only to approximate the shape of its time series by focusing on the normalized $N_a HHI_L$. After all, in this case, we do not need to estimate $N_a HHI_L$ precisely as both its mean and variance are irrelevant, allowing us to employ relatively aggregate employment data and simple assumptions regarding the relevant markets to compute it. We discuss the details of this computation in Section 3.4.4, which involves plugging estimates of firm-level employment data into $N_a HHI_L = N_a \sum_{i=1}^{N} (L_i/L)^2$.

3.3.3 Two-stage calibration procedure

Given our distributional assumption and the computed data moments, we quantify the model using a two-stage procedure. In this process, we parsimoniously assume that k is time-invariant.

First stage. For a given shape parameter $k \neq 0$, we compute $\underline{A}$ and $\overline{A}$ by matching aggregate TFP $\overline{A}\Omega$ and cost-weighted average markup μ to data, in a period-by-period basis. In Chapter 2, we showed that a unique solution for this calibration problem exists if and only if $\mu > 1$ and $k < 2/(\mu - 1)$. Under such conditions, given data on $\overline{A}\Omega$ and μ, one can find the unique solution for $\underline{A}$ and $\overline{A}$ from the following algorithm:

1. Given Equation (3.9), calculate $\tilde{A} \equiv \overline{A}/\underline{A}$ by solving numerically
$$\mu = \frac{1 - \mathrm{E}_a\left(\underline{A}/A\right)}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{A}/A)\right]} \tag{3.10}$$
which is obtained from (3.5) and (3.6) holding exactly as in the continuous model.

2. From Equation (3.6) holding exactly, compute $\underline{A} = \overline{A}\Omega/\mu$.

3. Given $\tilde{A} \equiv \overline{A}/\underline{A}$ and $\underline{A}$ from the previous steps, calculate $\overline{A} = \tilde{A} \times \underline{A}$.

Note computing $\tilde{A}$ from (3.10) only requires data on average markup μ. As a consequence, for a given $k \neq 0$, allocative efficiency Ω is just a function of μ, since, under (3.9), the calculation of Ω uses only $\tilde{A}$. Figure 3.1 plots this function Ω of μ for truncated Pareto distributions with $k = 3, 5, 9$ and an Uniform distribution $(k = -1)$. Several things are worth noting about it. First, Ω is strictly decreasing in $\mu = \frac{1-\alpha}{LS}$ and thus strictly increasing in LS. In particular, $\Omega \to 1^-$ when $\mu \to 1^+$. Intuitively, a lower average markup $\mu > 1$ indicates a less distorted economy, being associated with enhanced allocative efficiency Ω. Second, given a time series of μ, a lower k would imply a higher and less volatile estimated Ω.

Second stage. Given time series of $\overline{A}\Omega > 0$ and $\mu > 1$, from the first-stage algorithm one can obtain $\underline{A}$ and $\overline{A}$ in all periods under any $k < 2/(\max\{\mu\} - 1)$,

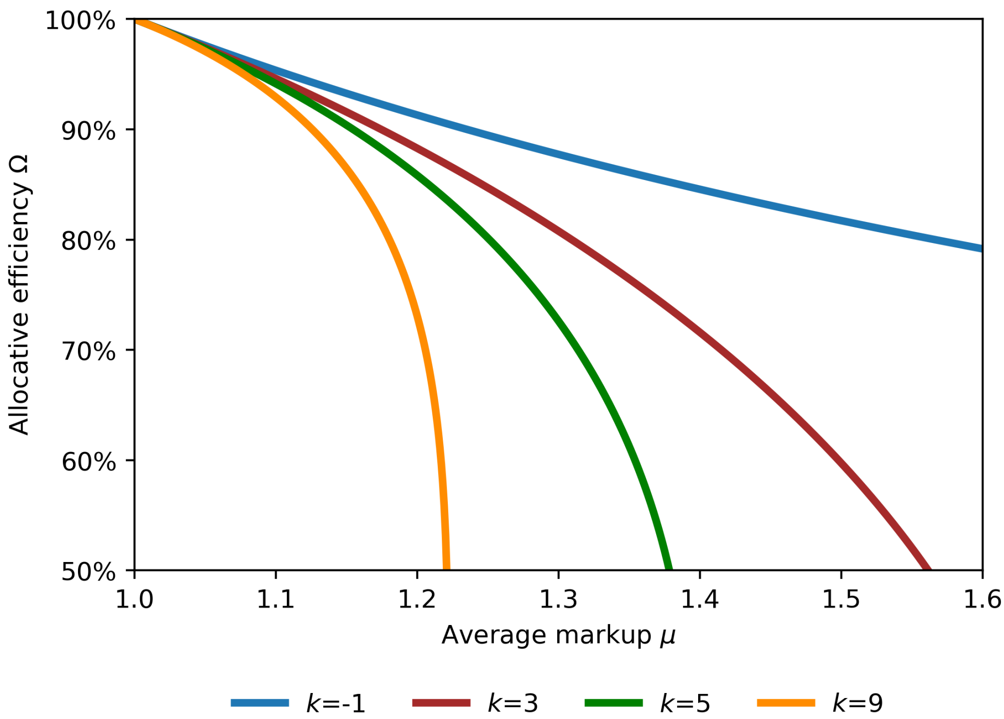


Figure 3.1 – Allocative efficiency vs. average markup.

$k \neq 0$, where $\max\{\mu\}$ denotes the maximum μ in its time series. And given $\underline{A}$ and $\overline{A}$, we can compute several variables in the model such as $N_a HHI_L$ from Equation (3.7). Under such conditions, we find the time-invariant $k \neq 0$ by searching the one that minimizes the Euclidean distance between the time series of $N_a HHI_L$ in the model and in the data when both series are normalized to have zero mean and unit variance.[4] Naturally, we limit the search to $k < 2/(\max\{\mu\} - 1)$. By analogous reasoning to previously employed for Ω, computing $N_a HHI_L$ requires only data on average markup μ for a given $k \neq 0$. As a result, the estimation of k is also independent of TFP data, implying that these data are not used to estimate Ω, not even through k. They are used only to pin down $\underline{A}$ and $\overline{A}$. Therefore, to decompose the TFP, we first estimate Ω using data on $N_a HHI_L$ and $\mu = \frac{1-\alpha}{LS}$, computing then $\overline{A}$ from the observed TFP $\overline{A}\Omega$. As a consequence, in this model, the residual of the production function is not the TFP $\overline{A}\Omega$ itself, but rather only its technology

[4] This is equivalent to search the k that minimizes the sum of the squared residuals of the normalized $N_a HHI_L$.

component $\overline{A}$, which is a cleaner residual as it is free of misallocation effects.

3.4 Data and parameters

3.4.1 Main data

We use an annual database, in which Y is the real GDP in 2010 prices, LS is the labor share of national income, and $\tilde{L}$ is the number of persons engaged, all from the Table of Resources and Uses of National Accounts produced by the Brazilian Institute of Geography and Statistics (IBGE). The labor share LS is calculated as the share of wages and social contributions in wages, social contributions, and gross operating surplus. Hence, self-employment income (mixed income) is allocated to labor and non-labor in the same proportions as the rest of the economy, following the preferred method of the Penn World Table 10.01 (Feenstra; Inklaar; Timmer, 2015). In Section 3.6.3, we evaluate the impact of using alternative labor share measures. Average annual hours worked by persons engaged h is from the Penn World Table 10.01, while we use Júnior and Cornelio (2020) for capital stock in 2010 prices $\tilde{K}$. To get K and L, we adjust respectively $\tilde{K}$ and $h\tilde{L}$ using the capacity utilization of manufacturing γ from the Getulio Vargas Foundation (FGV). This is accomplished through a flexible approach, considering the following empirical version of production function (3.4) for the year t:

$$Y_t = \overline{A}_t \Omega_t \tilde{K}_t^{\alpha} (h_t \tilde{L}_t)^{1-\alpha} \gamma_t^{\beta} \tag{3.11}$$

The underlying assumption is that capital and labor utilization is simply a rescaling of an observable measure of intensity in input usage, in line with the methodology of Basu, Fernald and Kimball (2006) and Basu et al. (2013). As argued in Basu et al. (2013, p.43), "[...] a cost-minimizing firm operates on all margins simultaneously, both observed and unobserved. As a result, changes in observed margins can proxy for unobserved utilization changes." Basu, Fernald and Kimball (2006) derive this result from a dynamic cost-optimizing firm problem, in which the firm is subject to adjustment costs and seeks to minimize the present discount value of its costs.

Note Equation (3.11) has two unknown parameters, α and β. In principle, both parameters could be estimated, but we find that (i) these estimates are highly dependent on the specific empirical strategy and estimator applied and (ii) α estimates are often implausibly high (or implausibly low if no capacity utilization adjustment is considered, i.e., if β is set to 0). As a result, we opt to initially calibrate the share parameter α before estimating β and, consequently, the TFP $\overline{A}_t\Omega_t$. This sequential approach mirrors the methodology employed by Basu, Fernald and Kimball (2006) and Basu et al. (2013).

3.4.2 Calibration of the share parameter

Since firms use inputs optimally, it is easy to show that α equals the cost share of capital, i.e., $\alpha = \frac{Kr}{Kr+Lw}$. Thus, we can calibrate α using cost share data, as we did in Chapter 2 for the US. Labor expenses Lw can be easily obtained from National Accounts. The main issue is to properly calculate capital expenses, distinguishing them from pure profit. To overcome this challenge, we follow Barkai (2020) and compute a required rate of return in the spirit of Hall and Jorgenson (1967), obtaining capital expenses by multiplying it by the nominal value of the capital stock. We do this for each type of capital disclosed by Júnior and Cornelio (2020): (i) residential structures, (ii) infrastructure, (iii) other structures, (iv) machinery and equipment, and (v) others.

Ignoring any tax treatment, the required rate of type s capital is the nominal weighted average cost of capital (WACC) minus the expected inflation of capital s plus the depreciation rate of capital s. Depreciation rates are from Júnior and Cornelio (2020). Expected inflations are the realized inflations of the capital goods computed using the Table of Resources and Uses of National Accounts. WACC data come from Santos (2020), in which Brazilian firms are supposed to finance investment using only (i) equity or (ii) debt with the Brazilian Development Bank (BNDES). The debt weight is given by BNDES lending for capital goods as a share of private sector gross capital formation. The debt cost is the BNDES lending rate for capital goods, while the equity cost is the Pre-DI swap 360 days plus the Equity Risk Premium (ERP) computed by Carvalho and Santos (2020).[5]

[5] "The Pre-DI swap contract traded at the BM&FBovespa, the Brazilian Stock Exchange, is an interest rate swap where one of the parties agrees to make pre-fixed interest payments in exchange for receiving floating interest payments based on the DI rate, whereas the other assumes a reverse position" (Carvalho; Santos, 2020, p.12).

These data allow us to compute the share of capital expenses on total expenses from 2002 to 2017 shown in Figure 3.2. Using this time series, we estimate an autoregressive process of order one and use its steady state to calibrate α. The idea behind this procedure is that the optimal use of factors is a better approximation over the long run, since actual factors use may significantly deviate from optimal static levels in the short run, especially for capital. We find $\alpha = 0.39$, higher than the standard calibration $\alpha = 1/3$. This result is robust to some changes in the calculation of the required rate of return. First, computing the equity cost using the Pre-DI swap 1800 days yields again $\alpha = 0.39$. Second, using the expected 12-month consumer inflation from the Focus survey instead of realized capital inflation does not change the estimate either. Third, giving null weight to debt cost and thus making the WACC equal to the equity cost increases the estimate only slightly to $\alpha = 0.40$. Fourth, we add corporate income tax following Barkai (2020). We find $\alpha = 0.38$ when we use 34% for the corporate (IRPJ/CSLL) tax rate and 100% for the capital cost recovery rate, which is the net present value of depreciation allowances for capital. Maintaining the same tax rate but setting this recovery rate to 70.71%, the OECD average level in 2021 according to the Tax Foundation, the estimate increases, but not by much, to $\alpha = 0.41$.

3.4.3 Estimation of the TFP

Assume $\Delta \ln(\overline{A}_t \Omega_t) = c + \xi_t$, where Δ is the first difference operator, $\mathrm{E}(\xi_t | I_{t-1}) = 0$, and I_{t-1} is the information set at $t-1$. Thus, from Equation (3.11),

$$\Delta \ln Y_t = c + \alpha \Delta \ln \tilde{K}_t + (1-\alpha) \Delta \ln(h_t \tilde{L}_t) + \beta \Delta \ln \gamma_t + \xi_t \qquad (3.12)$$

Setting $\alpha = 0.39$, we obtain the TFP by estimating c and β using Equation (3.12) and the moment conditions $\mathrm{E}\left(X_{t-1} \cdot \xi_t\right) = 0$, where X_{t-1} is a vector of instrumental variables such as $X_{t-1} \in I_{t-1}$. Essentially, this is an application, for a time series, of the simplified model of the dynamic panel literature (Blundell; Bond, 2000) presented in Ackerberg, Caves and Frazer (2015). Alternatively, we could have used the method of Olley and Pakes (1996) or its extensions Levinsohn and Petrin (2003) and Ackerberg, Caves and Frazer (2015). However, given $\Delta \ln(\overline{A}_t \Omega_t) = c + \xi_t$, these methods are equivalent to our own if all productivity shocks in (3.12) affect input

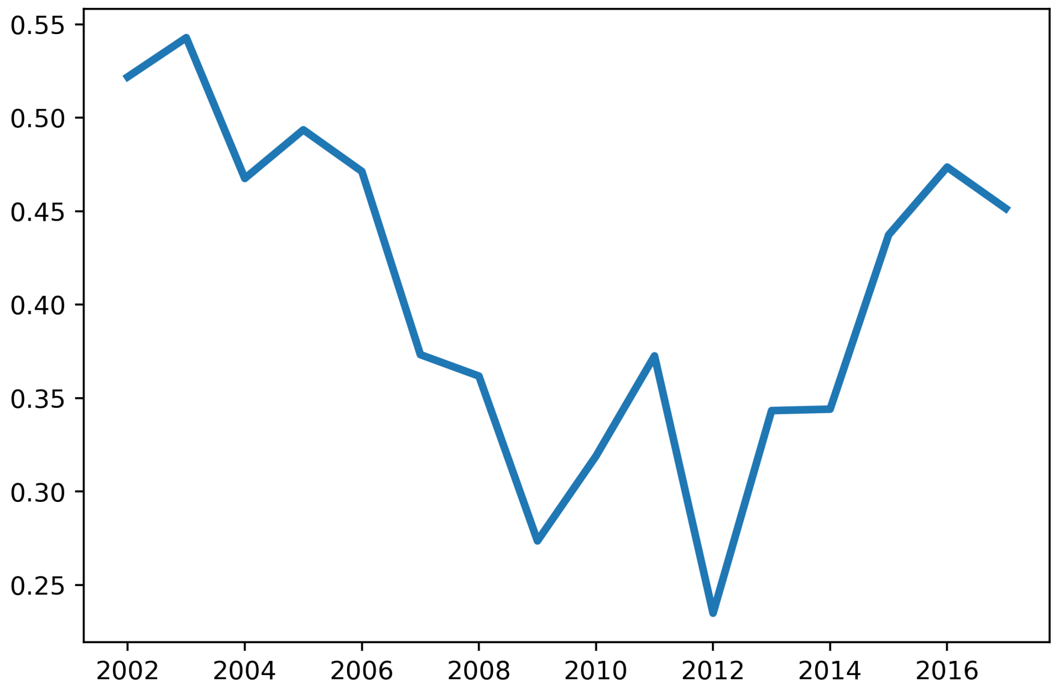


Figure 3.2 – Cost share of capital.

decisions since their first stage can be ignored in this case. And this assumption seems plausible here, as the inputs used in the production can be quickly adjusted by changing the utilization rates. Empirical evidence also seems to support it, as we do not find evidence of serial correlation in the error term.[6] Moreover, such methods require estimating a sufficiently flexible nonparametric function in their first stage (e.g., high-order polynomial), which would be highly problematic here due to the low sample size (in our current data set, we have 24 observations at most).

Our instruments X are a constant, $\ln \tilde{K}_t$, $\ln \tilde{K}_{t-1}$, $\ln(h_{t-1}\tilde{L}_{t-1})$, $\ln(h_{t-2}\tilde{L}_{t-2})$, $\ln \gamma_{t-1}$, $\ln \gamma_{t-2}$, $\ln Y_{t-1}$ and $\ln Y_{t-2}$. Under the assumption that all productivity shocks affect input decisions, these are the instruments suggested by Ackerberg,

[6] If $\ln(\overline{A}_t\Omega_t) = \omega_t + \epsilon_t$, where $\Delta\omega_t = c + \xi_t$ and ϵ_t denotes white-noise productivity shocks not considered by firms in input decisions, the error term in (3.12) would become $\xi_t + \Delta\epsilon_t \sim MA(1)$. Similar result apply under the more general assumption that $\omega_t = c + \rho\omega_{t-1} + \xi_t$, when one should "$\rho$-differentiate" (3.11) to get the expression analogous to (3.12), whose error term is $\xi_t + (\epsilon_t - \rho\epsilon_{t-1}) \sim MA(1)$ (Blundell; Bond, 2000).

Caves and Frazer (2015), expanded to include one additional lag of each to better account for the first differences included in (3.12). The instrumental variables estimator applied is the Two-Stage Least Squares (2SLS). For the sake of completeness, we also estimate (3.12) using Ordinary Least Squares (OLS).

Table 3.1 shows the estimates obtained using annual data between 1998 and 2019.[7] [8] They reject the no adjustment case ($\beta = 0$) and the usual practice of applying γ only over $\tilde{K}$ ($\beta = 0.39 = \alpha$). Indeed, the β estimates are much higher than $\alpha = 0.39$, which is consistent with labor hoarding. The impact of a higher β on the estimated TFP can be seen in Figure 3.3, which shows the TFP considering (i) no adjustment ($\beta = 0$), (ii) standard adjustment ($\beta = \alpha = 0.39$), (iii) OLS estimate ($\beta = 0.66$), and (iv) 2SLS estimate ($\beta = 0.74$). As can be seen, a higher β yields a smoother TFP, especially around the economic crisis of 2008 and 2015-2016. As seen in Section 3.3.3, the allocative efficiency Ω does not depend on TFP data in the Pareto case, implying a smoother TFP results in a smoother technology frontier $\overline{A}$, with no impact on Ω.

Table 3.1 – Estimates of production function (3.12) parameters for $\alpha = 0.39$

Parameter	OLS	2SLS
c	0.01***	0.01***
	(0.002)	(0.003)
β	0.656***	0.744***
	(0.089)	(0.095)

Parenthesis: Heteroskedasticity-robust standard error.
Significance: * (10%), ** (5%), *** (1%).
Sample (adjusted): 1998-2019 (22 observations) .

In what follows, we use the 2SLS β estimate as it is consistent under the assumptions considered here. In any case, this choice is not crucial for our empirical results since the TFPs computed using 2SLS and OLS β estimates are very similar, as can be seen in Figure 3.3.

[7] We have data from 1996 to 2019, but we lose the first two observations due to the existence of lagged variables.

[8] We also tested two generalized methods of moments (GMM) estimators described in (Hansen, 2020) that are efficient under heteroskedasticity: (i) two-step GMM, using 2SLS in the first step, and (ii) iterated GMM. The β estimates are only slightly higher (0.838 and 0.846, respectively), resulting in almost equivalent TFPs.

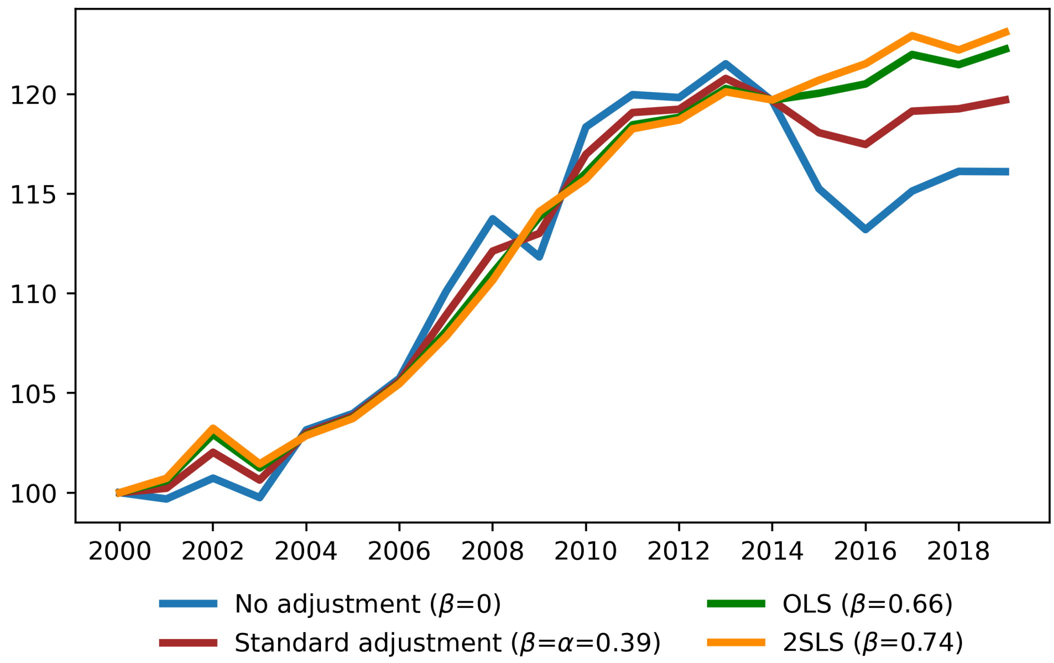


Figure 3.3 – TFP estimates for different values of β (2000=100).

3.4.4 Labor market data

We estimate $N_a HHI_L$ without any firm-level data, using two tables from the Central Register of Enterprises (CEMPRE) made available by the IBGE. The first one shows the total number of firms, total employment, and the number of firms and employment by nine size bins based on the number of employees.[9] The second one discloses the total number of firms, total employment, and employment share of the 4, 8, and 12 largest firms (in number of employees). Although CEMPRE covers all formal organizations, this second table is only available for formal corporate entities. Given that, we do not include data from other formal organizations, namely public administration and non-profit organizations, in the first table. This choice also aligns with the model utilized in this paper, where firms are profit-maximizing.

These two tables are available from 2006 to 2020 for all sectors (up to 3-digit level) from the National Classification of Economic Activities (CNAE) 2.0, which

[9] The bins are 0 to 4, 5 to 9, 10 to 19, 20 to 29, 30 to 49, 50 to 99, 100 to 249, 250 to 499, and 500 or more.

is a Brazilian classification derived from the ISIC Rev.4. Using this database, we follow five steps to compute $N_a HHI_L$.

Step 1 – addressing missing data. The IBGE does not disclose some employment information to avoid identifying the firms. To address it, for each 2-digit and 3-digit CNAE 2.0 sector and year, we do three imputations:

1. Use total employment from one table (if available) in the other (if missing).
2. Estimate missing employment data for a certain bin of firm size as the simple average of employment bin bounds multiplied by the number of firms.[10] [11]
3. If no missing data is left, we multiply all imputed employment in bins of firms' size by a scalar, ensuring that employment by bins adds to total employment. If only one missing data is left, it is backed out as a residual. If there is more than one missing data left, data from this sector in that year are discarded.

Naturally, the second imputation is not done for the last bin (500 or more employees) since we do not know its upper bound. Regarding the first bin (0 to 4 employees), even though it can be applied, the results may be distorted due to firms with no employees (for instance, if the number of such firms is disproportionately high). So, we consider two possibilities for the first bin, using or not using this second imputation procedure.

Step 2 – estimating the number of employees per firm. Denote the set of the lower bounds of firms' size bins from the first table by $b = \{500, 250, 100, ..., 0\}$, the k-th greatest element of b by b_k, and the number of firms with j or more employees by N_j. Given that, from the two tables already addressed for (some) missing data in step 1, we can get, for all available 2-digit and 3-digit CNAE 2.0 sectors in each year, the employment for the i largest formal enterprises, for $i \in I = \{4, 8, 12, N_{b_1}, N_{b_2}, ..., N_{b_9}\}$. Denote the j-th lowest element of I as i_j, with

[10] Thus, we are assuming that the average number of employees per firm in a given bin equals the simple average of employment bin bounds.

[11] For instance, if in a given sector and year there are 10 enterprises with 5 to 9 employees, we estimate its employment as $\frac{5+9}{2} \times 10 = 70$.

$j \in \{1, 2, ..., 12\}$. Then, for a given sector and year, we obtain the number of employees per firm by assuming that, for each $j \in \{1, 2, ..., 12\}$, the $(i_{j-1}+1)$-th to (i_j)-th largest firms have the same size, where $i_0 = 0$. In doing so, note that firms' size would be necessarily consistent with the bins' bounds of the first table.

Since the first bin (0 to 4 employees) includes firms with no employees, we also use two alternative methods that adjust the data by firms' size before applying the procedure just described. First, we exclude all first-bin data, such that $j \in \{1, 2, ..., 11\}$. Second, we maintain $j \in \{1, 2, ..., 12\}$, but replace $N_{b_9} = N_0$ by N_1, i.e., We exclude firms with no employees from the first bin. Since the IBGE does not disclose the number of such firms, we estimate it by assuming that firms with 1 to 4 employees have, on average, $\frac{1+4}{2} = 2.5$ employees. So, the new number of firms in the first bin is its total employment divided by 2.5.

Step 3 – computing desired results for each sector. Using the number of employees per firm from step 2, we compute, for all available 2-digit and 3-digit CNAE 2.0 sectors in each year, the number of active firms N_a and the Herfindahl–Hirschman concentration index HHI_L.

Step 4 – selecting sectors. We now select the sectors to be used in computing aggregate results, which should be representative of the entire economy and suitable for model calibration. We begin with all 3-digit CNAE 2.0 sectors since they are comprehensive, define the narrowest possible markets in data, and use the most granular information available in estimating the number of employees per firm, making it more precise.[12] Alternatively, we begin with only the 3-digit sectors that do not belong to a 2-digit sector that has a high share of non-enterprises formal organizations and/or we consider to be unsuitable to the model of Chapter 2 (e.g., utilities).[13]

[12] After all, this estimation relies on supposing that some firms are of the same size, and this restriction becomes weaker when using more granular data because, in this case, the number of firms assumed to have exactly the same size decreases.

[13] The 19 excluded 2-digit sectors are: 35 - Electricity, gas and other utilities, 36 - Water collection, treatment and distribution, 37 - Sewage and related activities, 38 - Collection, treatment and disposal of waste, and material recovery, 39 - Decontamination and other waste management services, 68 - Real estate activities, 72 - Scientific research and development,

We then adjust the initial sets of sectors to ensure the same sectoral coverage over time, avoiding distortions in the shape of the times series of the aggregate N_aHHI_L due to changes in the sectors included each year. We use two different adjustments. First, if a selected 3-digit sector is missing in step 3 for at least one year, we exclude all selected 3-digit sectors belonging to the corresponding 2-digit sector and, if its data are available for all years, add this 2-digit sector. Second, we simply exclude all selected 3-digit sectors with missing data in step 3.

Step 5 – aggregating desired results. Given the results obtained in step 3 for all the sectors selected in step 4, we follow three methods to obtain the aggregate results for each year. First, we assume there is only one national market, computing the aggregate number of active firms as the sum of the number of firms in each selected sector and the aggregate concentration index by plugging the number of employees of each firm in the economy directly into $HHI_L = \sum_{i=1}^{N} \left(\frac{L_i}{L}\right)^2$. Second, we compute N_aHHI_L for each selected sector and take their average. Third, we compute N_aHHI_L for each selected sector and take their median.

Table 3.2 summarizes all the methods to compute the desired results that we have discussed in the presentation of the five steps. We use all possible combinations of such methods, obtaining $2 \times 3 \times 2 \times 2 \times 3 = 72$ different time series of N_aHHI_L.

3.5 Results

Using the method discussed in Section 3.3 and the data and parameters of Section 3.4, we can obtain k, $\underline{A}$, and $\overline{A}$ for Brazil. We follow two steps. First, for a given time series of N_aHHI_L, we estimate the model from 2006 to 2019, finding

81 - Services for buildings and landscape activities, 84 - Public administration, defense and social security, 85 - Education, 86 - Human health care activities, 87 - Human health care activities integrated with social assistance, provided in collective and private residences, 88 - Social assistance services without accommodation, 90 - Artistic, creative and entertainment activities, 91 - Activities related to cultural and environmental heritage, 92 - Gambling and betting exploration activities, 93 - Sports, recreation and leisure activities, 94 - Activities of associative organizations, and 99 - International organizations and other extraterritorial institutions.

Table 3.2 – Alternative methods for computing $N_a HHI_L$

Step	Method	Method
1	Imputation in the first bin (0 to 4 employees)	(i) Average bin bounds × number of firms (ii) Backed out as a residual (if it is possible)*
2	Data from the first bin (0 to 4 employees)	(i) Included (ii) Included only for firms with employees* (iii) Excluded
4	Initial set of sectors	(i) All 3-digit sectors* (ii) Only selected 3-digit sectors
4	Sectors with missing data	(i) Replaced by the corresponding 2-digit sectors* (ii) Excluded
5	Aggregation method	(i) National market (ii) Sectors' average (iii) Sectors' median*

*Baseline methods.

the time-invariant k. Second, setting this k, we estimate all model parameters from 2000 to 2019 using the first-stage algorithm of the quantification method.

Figure 3.4 shows the $N_a HHI_L$ computed using the baseline methods highlighted in Table 3.2, when $k = 3.19$. In any case, choosing any other of the 72 time series of $N_a HHI_L$ would yield very similar results as all estimates of k are very close, between 3 and 3.3. It is worth mentioning that the results shown in Figures 3.4a, 3.4b, 3.4c, and 3.4d do not depend on TFP data. So, they are robust to other measures of labor and capital, including due to other adjustments by utilization. Only Figure 3.4e relies on TFP data.

Consistent with our quantification strategy, the model reproduces perfectly $\overline{A}\Omega$ and μ, but not $N_a HHI_L$ as we only seek to minimize the distance to this moment normalized. This can be seen in Figure 3.4a, in which the actual $N_a HHI_L$ is normalized to have the same average and standard deviation of the model time series between 2006 and 2019. In any case, the model fits the normalized data quite well, showing a correlation of 0.67 and capturing the upward trend seen in the data. It is also worth mentioning that the model series varies very little, always between 1.332 and 1.335.

Another way to gauge the model fit is to evaluate the number of active firms N_a. To compute the actual number of firms, we use data from Section 3.4.4 and the baseline methods highlighted in Table 3.2, but now for N_a.[14] To compute N_a in the model, we use Equation (3.3), with η replaced by η/q as in its continuous version, assuming η/q is time invariant. Analogously to N_aHHI_L, properly identifying the level of the actual N_a would require knowing the relevant market for each firm in the economy. Thus, we should evaluate the model fit only after normalizing the series. For instance, in Figure 3.4b, both series are divided by their 2006-2019 averages, which allows us to choose any feasible time-invariant adjusted elasticity η/q (e.g., $\eta/q = 1$) since changing it would simply multiply the model series by a constant. As can be seen, even though it is not a target moment, the model has a good fit as these series are highly correlated (coefficient of 0.85), with the model capturing the overall upward trend observed in data. The errors are concentrated around periods of economic crisis (2009 and 2015-2016) when the our static model is expected to behave poorly because, for instance, firms' use of the labor input is probably far from optimal and thus $\frac{1-\alpha}{LS}$ would not properly measure the cost-weighted average of firm-level markups μ.

We could also assess the concentration in the product market from N_aHHI_s. Unfortunately, we do not have data to compute it, but we can get it in the model from Equation (3.8). Figure 3.4c shows these model estimates, which point to an increase in product market concentration until 2004, a downward trend between 2005 and 2015, and a slight increase from 2016 to 2019. Note also that N_aHHI_s and N_aHHI_L behave very differently, being negatively correlated (-0.92) and with N_aHHI_s changing much more intensely.

Figure 3.4d presents the estimated allocative efficiency Ω. It shows an upward trend, reflecting the observed increase in the labor income share and, thus, the estimated decrease in the average markup. This is a sharp contrast with most developed countries, which over the last decades experienced decreasing labor share and increasing average markup (Calligaris; Criscuolo; Marcolin, 2018; Loecker; Eeckhout, 2018; Autor et al., 2020). Accordingly, studying the US, Baqaee and Farhi (2020) find that allocative efficiency, as measured by the distance from optimal allocation, deteriorated in 2015 relative to 1997. Moreover, Ω has essentially

[14] Hence, for each year, we get N_a for all selected sector and then take their median to obtain the aggregate value.

the same trends of N_aHHI_s, but with inverted signs (correlation of -0.9998): it decreases until 2004, increases from 2005 to 2015, and slightly decreases from 2016 onwards. Hence, N_aHHI_s seems to be a good proxy for Ω, with higher (lower) in N_aHHI_s suggesting lower (higher) Ω. Furthermore, Ω proved procyclical, with a 0.88 correlation with GDP.

Thus, similarly to the TFP $\overline{A}\Omega$, the allocative efficiency Ω increases during booms and decreases during busts, suggesting the TFP net of allocative efficiency effects, which is the technology frontier $\overline{A}$, should have more stable growth than the TFP.[15] This can be seen in Figure 3.4e, which plots $\overline{A}\Omega$ and $\overline{A}$ when both are adjusted to equal 100 in 2000. Indeed, using these adjusted series, we find a standard deviation of 8.39 for $\overline{A}\Omega$ and 6.01 for $\overline{A}$. Similarly, detrending both series using linear trends, we find a higher standard deviation for TFP (1.93 *versus* 1.54).

These results are easier to see in Table 3.3, which shows GDP Y, TFP $\overline{A}\Omega$, technology frontier $\overline{A}$, and allocative efficiency Ω average annual growth for the full sample (2000-2019) and some subperiods. We obtain the subperiods 2000-2003, 2003-2013, 2013-2016, and 2016-2019 by regressing the GDP growth rate from 1997 to 2019 on a constant, considering Newey-West robust standard errors and finding structural breaks using Bai-Perron tests. The results for these subperiods generally show that the TFP $\overline{A}\Omega$ is more volatile than the technology frontier $\overline{A}$. This becomes especially clear if one aggregates the two middle subperiods into one, when $\overline{A}$ has essentially the same growth in each remaining subperiod (0.89%, 0.9%, and 0.81% for 2000-2003, 2003-2016, and 2016-2019, respectively).

These results suggest that TFP growth cycles are mainly an allocative efficiency phenomenon. Indeed, the rapid productivity gains during 2003-2016 are essentially due to Ω, which grows 0.5% per year in the period *versus* -0.4% in 2000-2003 and 2016-2019. Focusing on the economic boom period of 2003-2013 would yield a similar conclusion when compared to 2000-2003 and 2016-2019 since, in this boom period, the technology frontier growth accelerated only slightly relative to these two subperiods. The conclusion changes only when this boom period is compared to 2013-2016 because, in this case, the higher TFP growth is mainly explained by faster technology improvements. However, this comparison is probably

[15] The correlation between TFP and GDP is 0.97.

not the most recommended one since Brazil experienced an intense economic crisis during 2013-2016, with GDP falling 7% between 2014 and 2016 (3.4% per year on average), when the static model employed here seems to behave poorly as suggested by the results of Figure 3.4b for N_a.

In short, the technology frontier grows much more steadily than the allocative efficiency, suggesting $\overline{A}$ is the most structural component of TFP. Therefore, since Ω could not increase or decrease indefinitely, the annual technology growth of 0.8-0.9% can be seen as the current structural, long-run, growth level of Brazilian TFP.

Table 3.3 – GDP, TFP $\overline{A}\Omega$, $\overline{A}$, and Ω annual growth for the baseline model, %

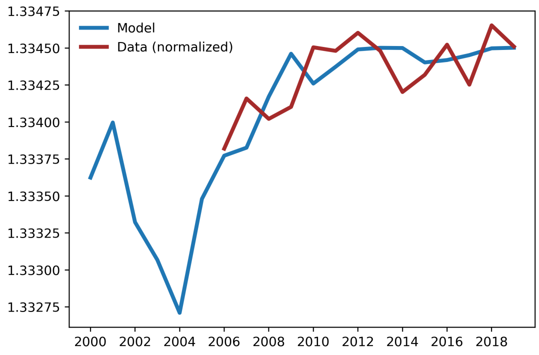
Variable	2000-2019	2000-2003	2003-2016			2016-2019
			2003-2016	2003-2013	2013-2016	
GDP Y	2.28	1.86	2.57	4.02	-2.12	1.44
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.89	0.89	0.9	1.16	0.04	0.81
Ω	0.21	-0.41	0.49	0.54	0.35	-0.37

3.6 Robustness

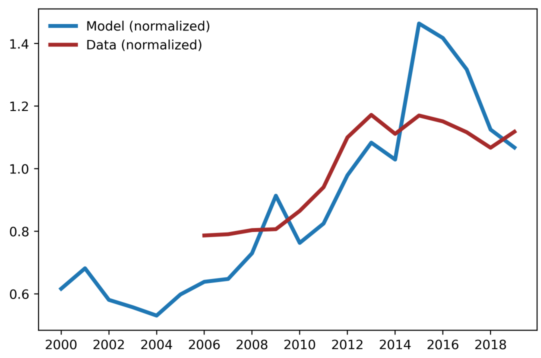
3.6.1 Shape parameter of the truncated Pareto distribution

From what we have seen in Section 3.3.3 from Figure 3.1, changes in LS seem to have a smaller impact on Ω for lower k. This fact suggests lowering k would yield a less volatile allocative efficiency Ω and thus a more volatile technology frontier $\overline{A}$. In other words, a lower k would yield more conservative results compared to standard growth-accounting exercises, in which all TFP growth is attributed to technology improvement. Indeed, we showed in Chapter 2 that $\Omega \to 1$ when $k \to -\infty$ and thus the model's residual $\overline{A}\Omega$ converges to the standard growth-accounting residual $\overline{A}$ in this limit case.

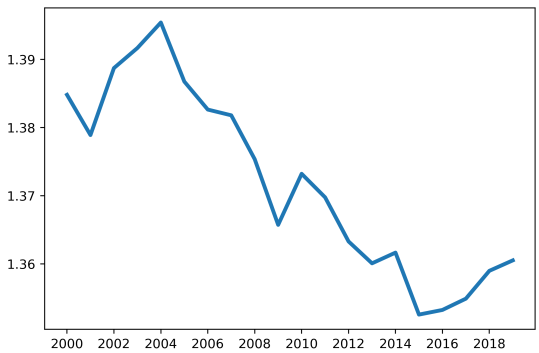
Given that, it would be useful to find a lower bound for k. As argued in Chapter 2, under the assumption that high-productivity firms are relatively scarce, this lower bound is $k = -1$. After all, the truncated Pareto density is



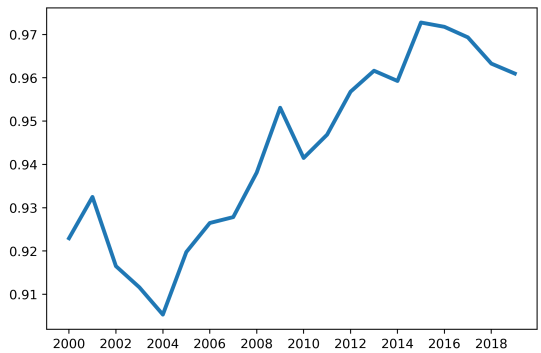
(a) $N_a HHI_L$



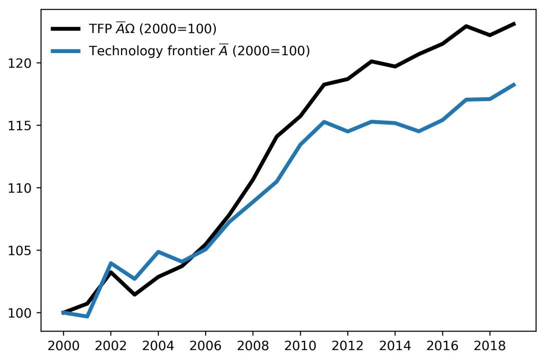
(b) N_a



(c) $N_a HHI_s$



(d) Allocative efficiency Ω



(e) TFP $\overline{A}\Omega$ and technology frontier $\overline{A}$

Figure 3.4 – Estimates for the baseline model.

downward sloping only for $k > -1$, $k \neq 0$, being upward sloping for $k < -1$, and constant for $k = -1$ (Uniform distribution). Besides being economically reasonable, this assumption is commonly adopted, often by imposing a non-truncated Pareto distribution for firms' productivity. In short, the most conservative reasonable TFP decomposition is obtained when the productivity is uniformly distributed.

Table 3.4 shows the new TFP decomposition. It is computed using the first-stage algorithm of the quantification method, which requires only macroeconomic data: $\mu = \frac{1-\alpha}{LS}$ is used to estimate Ω and then $\overline{A}$ is backed out as a residual in the TFP $\overline{A}\Omega$. The results show expected differences, particularly regarding the slightly higher variability of growth in $\overline{A}$ across the subperiods under $k = -1$. However, they are qualitatively similar, both supporting that TFP cycles are mainly due to allocative efficiency Ω. All in all, the main results seem to be robust to choosing conservatively $k = -1$.

3.6.2 Share parameter of the Cobb-Douglas production function

In Section 3.4.2, we use cost share data to get $\alpha = 0.39$. Here, we evaluate the robustness of our results to this choice. More precisely, we re-estimate the model parameters (including β) for (i) the lower standard $\alpha = 1/3$ and (ii) $\alpha = 0.41$, the highest calibration from Section 3.4.2 obtained by adding corporate income tax and setting the capital cost recovery rate to 2021 OECD average level. The new TFP growth decompositions are qualitatively similar, as shown in Table 3.4.

3.6.3 Labor share of national income

One major challenge in computing the labor income share is to gauge the labor income of self-employed workers since National Accounts typically disclose only the total income earned by such workers, known as mixed income. This issue is especially important in developing countries like Brazil, where a non-negligible portion of workers is self-employed.[16] Gollin (2002) proposes different ways to deal with that. One such way is to allocate mixed income to labor and non-labor in the

[16] In Brazil, about a quarter of employed people are self-employed according to 2012-2022 quarterly data from the National Household Sample Survey (PNAD) conducted by the IBGE.

same proportions as the rest of the economy. As discussed in Section 3.4.1, this is our baseline methodology and the preferred method of the Penn World Table 10.01. Alternatively, Gollin (2002) proposes allocating all mixed income to labor, which gives an upper bound for the labor income share LS. A lower bound for the labor income share can be obtained by not allocating any mixed income to labor, which is the naïve approach criticized by Gollin (2002). These three LS series can be seen in Figure 3.5. They are highly correlated but have different levels, as expected.

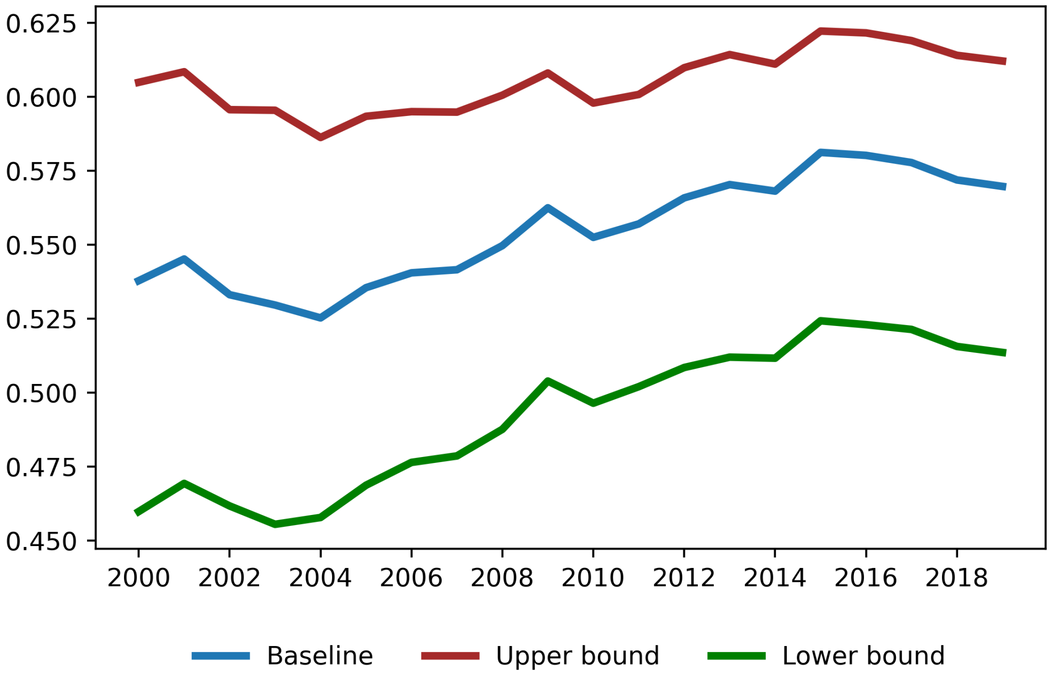


Figure 3.5 – Measures of labor income share.

As robustness exercises, we also apply our methodology using these two alternative labor income share data. For each new series, we initially recalculate the capital cost share, which would be higher for lower labor income, and re-calibrate α following the procedure of Section 3.4.2. Using the upper-bound (lower-bound) LS series, we find α equal to 0.37 (0.41). In any case, the average markup $\mu = \frac{1-\alpha}{LS}$ strictly decreases with the labor income share LS, which is not unexpected.[17]

[17] After all, $\alpha = \frac{Kr}{Kr+Lw} = \frac{KS}{KS+LS}$, where KS is the capital share of national income. Consequently, $\mu = \frac{1-\alpha}{LS} = \frac{1-\frac{KS}{KS+LS}}{LS} = \frac{1}{KS+LS}$.

Finally, for each alternative, we use the new α and the new LS series to re-estimate the other parameters (including β).

The results shown in Table 3.4 for the upper-bound LS series continue to support that TFP $\overline{A}\Omega$ cycles are mainly due to allocative efficiency Ω, but $\overline{A}$ growth becomes more volatile across the subperiods. This result is expected as this higher LS yields a smaller μ and, under such condition, the impact of changes of μ on allocative efficiency Ω seems to be smaller for $k > -1$, $k \neq 0$, as can be seen in Figure 3.1.

By analogous reasoning and given that labor share is highly correlated with GDP (coefficient of 0.88), one should expect a more cyclical allocative efficiency Ω for the lower-bound LS series. The results shown in Table 3.4 are consistent with this prediction. Indeed, in this case, Ω is so cyclical that the residual $\overline{A}$ becomes much more volatile and, more importantly, countercyclical in some periods. For instance, between 2000-2003 and 2003-2016, the growth of the technology frontier $\overline{A}$ decreased while the GDP and TFP accelerated, which does not seem to be intuitive. This suggests this naïve way of computing the labor income share is inappropriate for a developing economy like Brazil, as advocated by Gollin (2002).

In short, the results change qualitatively only under the unreasonable assumption that no mixed income is due to labor.

3.6.4 Zipf's law calibration

Until now, we have always quantified the model assuming firm productivity A is Pareto distributed. Although it is relatively standard in the literature, it is hard to evaluate the reasonably of this choice as firm productivity is hardly observable. What can be easily observed is the distribution of firm size, which usually follows a distributional power law known as Zipf's law.[18] Formally, under this "law", $P(S \geq s) = (\underline{s}/s)^k$ for some measure of firm size $S \geq \underline{s} > 0$, with $k \approx 1$. Empirical evidence supporting this claim has been found for several different countries with firm size measured by the number of employees, sales, income, total

[18] Power laws emerge in various domains within economics and finance. See Gabaix (2009), Gabaix (2016) for a review.

assets, and equity plus debt (Okuyama; Takayasu; Takayasu, 1999; Axtell, 2001; Fujiwara et al., 2004; Luttmer, 2007; Gabaix; Landier, 2008; Giovanni; Levchenko; Ranciere, 2011; Giovanni; Levchenko, 2013). Specifically to Brazil, Silva et al. (2018) find support for Zipf's law among the 1,000 largest firms by net revenue in 2015. Moreover, in Chapter 4, we evaluate the distribution of firm size by the number of employees using CEMPRE data. Remarkably, we find Zipf's law provides a very good, although not perfect, approximation to data for each year between 1996 and 2020 at the economy-wide level and also for agriculture, industry, and services alone.

Given that scenario, instead of choosing a density for A, we now impose the distribution of firms' market share $s(A)$.[19] As before, we choose to use a continuous distribution, consistent with the continuous model of Chapter 2. However, we cannot assume the market share is non-truncated Pareto distributed with shape parameter $k \approx 1$, when Zipf's law would hold over the entire support. On the one hand, we should use a truncated distribution since the market share is bounded. On the other hand, since $s(\underline{A}) = 0$, the support of the market share distribution is not strictly greater than 0. Under such circumstances, we assume the market share $s(A)$ follows a truncated Lomax distribution, a special case of the truncated Pareto distribution type II. More precisely, we assume $\tilde{s}(A) \equiv 1 - \underline{A}/A = \frac{s(A)}{\eta/q} \in \left[0, \tilde{s}(\overline{A})\right] = \left[0, 1 - \tilde{A}^{-1}\right]$ is truncated Lomax distributed with shape parameter $k \neq 0$ and scale parameter $\lambda > 0$ or, equivalently, $(\tilde{s}(A) + \lambda) \in [\lambda, \tilde{s}(\overline{A}) + \lambda]$ is truncated Pareto distributed with shape parameter $k \neq 0$. Therefore, when $\lambda \approx 0$, the firm market share is approximately truncated Pareto distributed. If we further consider $k \approx 1$, the distribution aligns to Zipf's law predictions, particularly away from support's bounds (see Appendix 3.C.1 for a discussion).

We left the details to Appendix 3.C, but we essentially follow the same two stages presented in Section 3.3.3 to quantify the model. In the first stage, given k and λ, we gauge $\underline{A}$ and $\overline{A}$ by matching aggregate TFP $\overline{A}\Omega$ and average markup μ to data. In the second stage, we set $k \approx 1$ and search the time-invariant λ that minimizes the distance to the normalized $N_a HHI_L$. Similarly to the baseline Pareto

[19] In monopolistic competition models like Melitz (2003), supposing firm productivity is Pareto distributed delivers a distributional power law in firm size (Giovanni; Levchenko; Ranciere, 2011). This does not hold for the Cournot model of Chapter 2.

case, TFP data are not required to estimate Ω, not even through λ. They are used only to pin down $\underline{A}$ and $\overline{A}$. Hence, as before, the residual of this model is not the TFP $\overline{A}\Omega$ itself, but rather only its technology component $\overline{A}$.

Table 3.4 shows the TFP decomposition under this new distribution for $k = 0.5, 1, 1.5$. The results are practically identical to the baseline model for all these three shape parameters. The estimated scale parameter λ is 0.55, 0.76, and 0.97 for k equal to 0.5, 1, and 1.5, respectively. Thus, the choice of k does not seem crucial since changing the assigned k alters the calibrated λ in the second stage, leaving the main model results basically unchanged.[20]

3.7 Conclusion

This paper employs the Cournot model of Chapter 2 to decompose the Brazilian TFP between 2000 and 2019. We find an overall improvement in allocative efficiency, reflecting the observed increase in the labor income share and, thus, the estimated decrease in the average markup. We also find that TFP cycles are essentially due to allocative efficiency, with the economic boom in the mid-2000s being primarily attributed to efficiency gains. The technology frontier grows much more steadily, suggesting this reflects the structural characteristics of the economy. Therefore, since allocative improvements could not occur indefinitely, the annual technology growth found, around 0.8-0.9%, can be seen as the current structural, long-run, growth level of Brazilian TFP.

The main conclusions are robust to using conservatively a Uniform distribution, when the model quantification requires *only* macroeconomic data. They are also robust to considering the standard $\alpha = 1/3$ or $\alpha = 0.41$ instead of our baseline calibration of $\alpha = 0.39$. Using a higher labor income share, in which all mixed income is allocated to labor, does not alter our main conclusions either. Finally, the results remain unchanged when we follow a different strategy, assuming the distribution of firm market share is consistent with Zipf's law instead of imposing a firm productivity distribution.

[20] When seeking both $k \neq 0$ and $\lambda > 0$ in the second stage, we find very high $k \approx 212$ and $\lambda \approx 90$, but the TFP decomposition is, once more, practically identical to the baseline model, confirming this intuition.

Table 3.4 – TFP $\overline{A}\Omega$, $\overline{A}$, and Ω annual growth, %

Variable	2000-2019	2000-2003	2003-2016			2016-2019
			2003-2016	2003-2013	2013-2016	
Baseline model						
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.89	0.89	0.9	1.16	0.04	0.81
Ω	0.21	-0.41	0.49	0.54	0.35	-0.37
Robustness #1: Uniform distribution ($k = -1$)						
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.95	0.74	1.04	1.33	0.1	0.74
Ω	0.15	-0.26	0.35	0.37	0.29	-0.3
Robustness #2: $\alpha = 1/3$						
TFP $\overline{A}\Omega$	1.13	0.49	1.45	1.75	0.47	0.36
$\overline{A}$	0.81	1.14	0.71	0.93	-0.04	0.9
Ω	0.32	-0.64	0.74	0.81	0.5	-0.53
Robustness #3: $\alpha = 0.41$						
TFP $\overline{A}\Omega$	1.09	0.47	1.37	1.68	0.35	0.46
$\overline{A}$	0.9	0.83	0.95	1.22	0.05	0.79
Ω	0.18	-0.35	0.42	0.46	0.3	-0.32
Robustness #4: upper-bound LS series						
TFP $\overline{A}\Omega$	1.11	0.48	1.42	1.72	0.41	0.41
$\overline{A}$	1.07	0.79	1.23	1.54	0.2	0.68
Ω	0.03	-0.3	0.18	0.18	0.21	-0.27
Robustness #5: lower-bound LS series						
TFP $\overline{A}\Omega$	1.09	0.47	1.37	1.68	0.35	0.46
$\overline{A}$	0.47	0.92	0.27	0.41	-0.2	0.94
Ω	0.61	-0.44	1.11	1.27	0.56	-0.47
Robustness #6: Zipf's law calibration ($k = 0.5$)						
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.89	0.87	0.91	1.18	0.04	0.81
Ω	0.21	-0.39	0.48	0.52	0.35	-0.37
Robustness #7: Zipf's law calibration ($k = 1$)						
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.89	0.87	0.91	1.18	0.04	0.81
Ω	0.21	-0.39	0.48	0.52	0.35	-0.37
Robustness #8: Zipf's law calibration ($k = 1.5$)						
TFP $\overline{A}\Omega$	1.1	0.48	1.4	1.7	0.39	0.43
$\overline{A}$	0.89	0.88	0.91	1.18	0.04	0.81
Ω	0.21	-0.39	0.48	0.52	0.35	-0.37

Appendix of Chapter 3

3.A Derivation of two concentration measures

Relying on the results of Chapter 2, we derive expressions for two concentration measures. First, the Herfindahl–Hirschman index (HHI) of the labor market $HHI_L \equiv \sum_{i=1}^{N}(L_i/L)^2$, where L_i is the labor employed by firm $i \in \{1, 2, ..., N\}$ and $L \equiv \sum_{i=1}^{N} L_i$. Second, the HHI of the product market $HHI_s \equiv \sum_{i=1}^{N} s(A_i)^2$, with $s(A_i)$ being the market share of an active firm with productivity A_i. We start with the baseline model, which we employ in the main text, but we also consider the two model extensions presented in the previous chapter.

3.A.1 Baseline model

In Chapter 2, we derived the following results for the discrete model:

$$\theta_i \equiv \theta_{Li} = \theta_{Ki} = \overline{A}\Omega \frac{s(A_i)}{A_i} \tag{3.A.1}$$

$$s(A_i) \approx \eta \left(1 - \underline{A}/A_i\right) \tag{3.2}$$

$$\eta \approx \frac{1}{N_a \left[1 - \mathrm{E}_a\left(\underline{A}/A\right)\right]} \tag{3.3}$$

$$\Omega \approx \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{A}/A)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{A}/A)\right]} \tag{3.5}$$

$$\mu \approx \frac{\overline{A}\Omega}{\underline{A}} \tag{3.6}$$

where $\theta_{Li} \equiv L_i/L$ and $\theta_{Ki} \equiv K_i/K$, with K_i being the capital stock of firm i and $K \equiv \sum_{i=1}^{N} K_i$. Aggregate TFP is denoted by $\overline{A}\Omega$, with Ω being a measure of allocative efficiency and $\overline{A}$ the productivity of the most efficient firm. Moreover, $\underline{A}$ is the lowest productivity level among active firms, η is the absolute value of the price elasticity of demand, N_a is the number of active firms, and $\mathrm{E}_a\left(h(A)\right)$ is the expected value of a function h over active firms under the empirical distribution.

Finally, being μ_i the markup of firm i, w the wage, and r the rental cost of capital, $\mu \equiv \sum_{i=1}^{N}\left(\frac{L_i w + K_i r}{Lw + Kr}\right)\mu_i$ is the cost-weighted average of firm-level markups.

Using these results, let us derive expressions for the two measures of concentration presented earlier. First, the HHI of the labor market $HHI_L \equiv \sum_{i=1}^{N}\theta_{Li}^2$. Plugging (3.A.1) into this definition and using (3.2) and (3.3),

$$\begin{aligned}
HHI_L =& \sum_{i=1}^{N}\left[\overline{A}\Omega\frac{s(A_i)}{A_i}\right]^2 \approx \left[\frac{\Omega(\overline{A}/\underline{A})}{1-\mathrm{E}_a\left(\underline{A}/A\right)}\right]^2 \frac{\mathrm{E}_a\left[(\underline{A}/A)^2\left(1-\underline{A}/A\right)^2\right]}{N_a} \\
N_a HHI_L \approx& \Omega^2 \frac{\mathrm{E}_a\left[(\underline{A}/A)^2\left(1-\underline{A}/A\right)^2\right]}{\left\{\mathrm{E}_a\left[(\underline{A}/\overline{A})\left(1-\underline{A}/A\right)\right]\right\}^2} \\
N_a HHI_L \approx& \frac{\mathrm{E}_a\left[(\underline{A}/A)^2\left(1-\underline{A}/A\right)^2\right]}{\left\{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]\right\}^2}
\end{aligned} \tag{3.7}$$

where in the last line we use (3.5). Second, the HHI of the product market $HHI_s \equiv \sum_{i=1}^{N} s(A_i)^2$. Plugging (3.2) and (3.3) into it, one gets

$$N_a HHI_s \approx \frac{\mathrm{E}_a\left[\left(1-\underline{A}/A\right)^2\right]}{\left[\mathrm{E}_a\left(1-\underline{A}/A\right)\right]^2} \approx \frac{1-1/\mu}{1-\mathrm{E}_a\left(\underline{A}/A\right)} \tag{3.8}$$

where, in the last part, we use (3.5) and (3.6) since, from these equations, it is easy to see that

$$\begin{aligned}
1-\frac{1}{\mu} \approx& 1-\frac{\underline{A}}{\overline{A}\Omega} \approx 1-\frac{\mathrm{E}_a\left[(\underline{A}/A)(1-\underline{A}/A)\right]}{1-\mathrm{E}_a\left(\underline{A}/A\right)} \\
1-\frac{1}{\mu} \approx& \frac{1-2\mathrm{E}_a\left(\underline{A}/A\right)+\mathrm{E}_a\left[(\underline{A}/A)^2\right]}{1-\mathrm{E}_a\left(\underline{A}/A\right)} = \frac{\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]}{1-\mathrm{E}_a\left(\underline{A}/A\right)}
\end{aligned} \tag{3.A.2}$$

In another version of the model, we considered a continuum of firms indexed by $i \in [0, N]$. In this case, the results shown at the beginning of this section would still be valid, but now holding *exactly*, if one replaces (i) sums by integrals (e.g., $K \equiv \int_0^N K_i di$) and (ii) η by η/q, with $q \in (0,1]$ being the marginal effect of increasing a firm's output on aggregate output considered by each firm. Given that, it is easy to see that Equations (3.7) and (3.8) are also valid for the continuous model, but in this case holding *exactly*. Naturally, the measures of concentration should defined accordingly, with $HHI_L \equiv \int_0^N \theta_{Li}^2 di$ and $HHI_s \equiv \int_0^N s(A_i)^2 di$.

3.A.2 Model extensions

In Chapter 2, we discussed two model extensions. The first extension goes beyond the Cobb-Douglas production function, considering an arbitrary well-behaved production function with constant returns to scale, Hicks-neutral productivity shifter, and M factors of production. In this case, the baseline Equations (3.A.1), (3.2), (3.3), (3.5), and (3.6) are still valid, implying (3.7) and (3.8) continue to hold (exactly for a continuum of firms).[21] The second extension adds firm-specific wedges as a new source of firm heterogeneity, considering firm-specific tax rate over revenue $\tau_i = \tau(A_i)$. In this case, for a discrete number of firms,

$$\theta_i \equiv \theta_{Li} = \theta_{Ki} = \overline{A}\Omega \frac{s(B_i)}{A_i} \tag{3.A.1”}$$

$$s(B_i) \approx \eta \left(1 - \underline{B}/B_i\right) \tag{3.2”}$$

$$\eta \approx \frac{1}{N_a \left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]} \tag{3.3”}$$

$$\Omega \approx \frac{\mathrm{E}_a\left[(\underline{A}/\overline{A})(1 - \underline{B}/B)\right]}{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{B}/B)\right]} \tag{3.5”}$$

$$\mu \approx \frac{\overline{A}\Omega}{\underline{B}}\left(1 - \tilde{\tau}\right) \tag{3.6”}$$

where $s(B_i)$ is the market share of a firm with adjusted productivity $B_i \equiv A_i\left(1 - \tau(A_i)\right)$, $\underline{B}$ is the lowest level of the adjusted productivity B_i among active firms, and $\tilde{\tau} \equiv \sum_{i=1}^{N} s(B_i)\tau(A_i)$ is the sales-weighted average tax rate. As a result, plugging (3.A.1”) into $HHI_L \equiv \sum_{i=1}^{N} \theta_{Li}^2$ and using (3.2”) and (3.3”),

$$N_a HHI_L \approx (\overline{A}\Omega)^2 \frac{E_a\left[\left((1/A) - \underline{B}/(A \times B)\right)^2\right]}{\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]^2} = \left(\frac{\overline{A}\Omega}{\underline{A}}\right)^2 \frac{E_a\left[(\underline{A}/A)^2(1 - \underline{B}/B)^2\right]}{\left[1 - \mathrm{E}_a\left(\underline{B}/B\right)\right]^2}$$

$$N_a HHI_L \approx \frac{E_a\left[(\underline{A}/A)^2(1 - \underline{B}/B)^2\right]}{\left\{\mathrm{E}_a\left[(\underline{A}/A)(1 - \underline{B}/B)\right]\right\}^2} \tag{3.7”}$$

where in the last line we use (3.5”). Finally, plugging (3.2”) and (3.3”) into $HHI_s \equiv \sum_{i=1}^{N} s_i^2$,

$$N_a HHI_s \approx \frac{\mathrm{E}_a\left[(1 - \underline{B}/B)^2\right]}{\left[\mathrm{E}_a\left(1 - \underline{B}/B\right)\right]^2} \tag{3.8”}$$

[21] Naturally, for HHI_L to make sense, labor should be one of the factors of production.

By analogous reasoning to that employed previously for the baseline model, Equations (3.7") and (3.8") would be *exactly* valid with a continuum of firms. As before, in this case, one should define $HHI_L \equiv \int_0^N \theta_{Li}^2 di$ and $HHI_s \equiv \int_0^N s(A_i)^2 di$.

3.B Conditions for the first-stage algorithm to work properly

In Chapter 2, we established necessary and sufficient conditions for the first-stage algorithm to work properly, achieving an exact match of both target moments. Initially, we derived some general results by examining an arbitrary truncated distribution. The only requirement is that this distribution can be expressed as a truncation of another density from above. These general findings provide a framework for establishing conditions applicable to any such distribution. In the previous chapter, we used this framework to evaluate the case in which firms' productivity is truncated Pareto distributed. Here, we apply it for a truncated Lomax distribution of firms' market share, showing in which conditions (3.10) implicitly defines $\tilde{A} \equiv \overline{A}/\underline{A}$ as a well-defined function of μ.

In the following, consider μ as given in (3.10) or, equivalently,

$$\mu = \frac{\mathrm{E}_a\left(1 - \underline{A}/A\right)}{\mathrm{E}_a\left(1 - \underline{A}/A\right) - \mathrm{E}_a\left[(1 - \underline{A}/A)^2\right]} \tag{3.10}$$

where $\mathrm{E}_a\left(h(A)\right) \equiv \mathrm{E}\left(h(A)|\underline{A} \leq A \leq \overline{A}\right)$ for any function h.

3.B.1 Arbitrary distribution

Assume $A \in [\underline{\underline{A}}, \overline{A}]$ is a continuous variable, $0 < \underline{\underline{A}} < \overline{A} < +\infty$, whose density and cumulative distribution function are g and G, respectively. Let $\tilde{g}(A) \equiv \frac{g(A)}{1-G(\underline{A})} > 0$, $\underline{A} \in (\underline{\underline{A}}, \overline{A})$, the density function of $A \in [\underline{A}, \overline{A}]$, with cumulative distribution $\tilde{G}$. Let $\hat{g}$ be another density of A, but defined over the support $A \in [\underline{A}, A_h]$, $A_h > \overline{A}$, possibly with $A_h \to +\infty$. This density does not depend on $\overline{A}$ and has cumulative distribution function $\hat{G}$. Moreover, it satisfies $\tilde{g}(A) = \frac{\hat{g}(A)}{\hat{G}(\overline{A})}$, meaning $\tilde{g}$ is a truncation of $\hat{g}$ from above.

Proposition 3.B.1 *$\overline{A}$, $\overline{A} > \underline{A}$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in \left(1, \lim_{\overline{A}\to+\infty} \mu\right)$.*

Proof. See Chapter 2. ■

3.B.2 Lomax distribution of firm market share

Let $\tilde{s}(A) \equiv 1 - \underline{A}/A$ for $A \in [\underline{A}, \overline{A}] \in (0, +\infty)$, with $\tilde{A} \equiv \overline{A}/\underline{A} > 1$. Assume $\tilde{s}(A) \in [0, \tilde{s}(\overline{A})] = [0, 1 - \tilde{A}^{-1}]$ is truncated Lomax distributed with shape parameter $k \neq 0$ and scale parameter $\lambda > 0$ or, equivalently, $(\tilde{s}(A) + \lambda) \in [\lambda, \tilde{s}(\overline{A}) + \lambda]$ is truncated Pareto distributed with the same parameter $k \neq 0$. Hence, the density of $\tilde{s}(A)$ is $\tilde{g}_s(\tilde{s}(A)) = k\left[\frac{\lambda^k(\tilde{s}(\overline{A})+\lambda)^k}{(\tilde{s}(\overline{A})+\lambda)^k-\lambda^k}\right](\tilde{s}(A)+\lambda)^{-k-1} = k\left(\frac{\lambda^k S^k}{S^k-1}\right)(\tilde{s}(A)+\lambda)^{-k-1}$, where $S \equiv \frac{\tilde{s}(\overline{A})}{\lambda} + 1$ is a function of $\tilde{A}$ and λ. Denote by $\tilde{G}_s$ the cumulative distribution function of $\tilde{s}(A)$. In the following, let $\overline{\mu} \equiv \lim_{\overline{A}\to+\infty} \mu$.

Proposition 3.B.2 *For $k \neq 0$, $\lambda > 0$, and $j \in \mathbb{N} \setminus \{0\}$,*

$$\mathrm{E}_a\left[(\tilde{s}(A) + \lambda)^j\right] = \begin{cases} \frac{k\lambda^j}{j-k}\left(\frac{S^j - S^k}{S^k-1}\right) & \text{, if } j \neq k \\ k\lambda^k\left(\frac{S^k \ln S}{S^k-1}\right) & \text{, if } j = k \end{cases}$$

Proof. Let $k \neq 0$, $\lambda > 0$, and $j \in \mathbb{N} \setminus \{0\}$. From the truncated Lomax density, if $j \neq k$,

$$\mathrm{E}_a\left[(\tilde{s}(A) + \lambda)^j\right] = k\left(\frac{\lambda^k S^k}{S^k - 1}\right)\int_0^{\tilde{s}(\overline{A})}(s+\lambda)^{j-k-1}ds$$

$$\mathrm{E}_a\left[(\tilde{s}(A) + \lambda)^j\right] = k\left(\frac{\lambda^k S^k}{S^k - 1}\right)\left[\frac{(\tilde{s}(\overline{A})+\lambda)^{j-k} - \lambda^{j-k}}{j-k}\right]$$

$$\mathrm{E}_a\left[(\tilde{s}(A) + \lambda)^j\right] = k\left(\frac{\lambda^k S^k}{S^k - 1}\right)\left[\frac{S^{j-k} - 1}{(j-k)\lambda^{k-j}}\right] = \frac{k\lambda^j}{j-k}\left(\frac{S^j - S^k}{S^k - 1}\right)$$

If $j = k$, $\mathrm{E}_a\left[(\tilde{s}(A) + \lambda)^j\right] = k\left(\frac{\lambda^k S^k}{S^k-1}\right)\left[\ln(\tilde{s}(\overline{A}) + \lambda) - \ln(\lambda)\right] = k\lambda^k\left(\frac{S^k \ln S}{S^k-1}\right)$. ■

Proposition 3.B.3 *For $\lambda > 0$,*

$$\mu = \begin{cases} \frac{k(S-1)-(S^k-1)}{\left(1+\frac{2\lambda}{2-k}\right)\left[k(S-1)-(S^k-1)\right]-\frac{\lambda k(1-k)}{2-k}(S-1)^2} & \text{, if } k \neq 0,1,2 \\ \frac{S\ln S-(S-1)}{(1+2\lambda)[S\ln S-(S-1)]-\lambda(S-1)^2} & \text{, if } k = 1 \\ \frac{(S-1)^2}{2(1+2\lambda)S(S-1)-2\lambda S^2\ln S-(1+\lambda)(S^2-1)} & \text{, if } k = 2 \end{cases}$$

Proof. Let $\lambda > 0$. Initially, note Equation (3.10) can be rewritten as

$$\mu = \frac{\mathrm{E}_a\left(1-\underline{A}/A\right)}{\mathrm{E}_a\left(1-\underline{A}/A\right)-\mathrm{E}_a\left[(1-\underline{A}/A)^2\right]} = \frac{\mathrm{E}_a\left(\tilde{s}(A)\right)}{\mathrm{E}_a\left(\tilde{s}(A)\right)-\mathrm{E}_a\left[\tilde{s}(A)^2\right]}$$

$$\mu = \frac{\mathrm{E}_a\left(\tilde{s}(A)+\lambda\right)-\lambda}{\mathrm{E}_a\left(\tilde{s}(A)\right)-\mathrm{E}_a\left[\left(\tilde{s}(A)+\lambda\right)^2\right]+\lambda^2+2\lambda\mathrm{E}_a\left[\tilde{s}(A)\right]}$$

$$\mu = \frac{\mathrm{E}_a\left(\tilde{s}(A)+\lambda\right)-\lambda}{(1+2\lambda)\mathrm{E}_a\left(\tilde{s}(A)+\lambda\right)-\lambda(1+2\lambda)-\mathrm{E}_a\left[\left(\tilde{s}(A)+\lambda\right)^2\right]+\lambda^2}$$

$$\mu = \frac{\mathrm{E}_a\left(\tilde{s}(A)+\lambda\right)-\lambda}{(1+2\lambda)\mathrm{E}_a\left(\tilde{s}(A)+\lambda\right)-\mathrm{E}_a\left[\left(\tilde{s}(A)+\lambda\right)^2\right]-\lambda(1+\lambda)}$$

since $\tilde{s}(A) \equiv 1-\underline{A}/A$, where $\mathrm{E}_a\left[h(\tilde{s}(A))\right] = \int_{\tilde{s}(\underline{A})}^{\tilde{s}(\overline{A})} h(s)\tilde{g}_s(s)ds$ for a function h.

Therefore, if $k = 1$, using Proposition 3.B.2 one can see that

$$\mu = \frac{\lambda\left(\frac{S\ln S}{S-1}\right)-\lambda}{(1+2\lambda)\lambda\left(\frac{S\ln S}{S-1}\right)-\lambda^2\left(\frac{S^2-S}{S-1}\right)-\lambda(1+\lambda)}$$

$$\mu = \frac{S\ln S-(S-1)}{(1+2\lambda)S\ln S-\lambda S(S-1)-(1+2\lambda-\lambda)(S-1)}$$

$$\mu = \frac{S\ln S-(S-1)}{(1+2\lambda)\left[S\ln S-(S-1)\right]-\lambda(S-1)^2}$$

If $k = 2$, from Proposition 3.B.2,

$$\mu = \frac{2\lambda\left(\frac{S^2-S}{S^2-1}\right)-\lambda}{(1+2\lambda)2\lambda\left(\frac{S^2-S}{S^2-1}\right)-2\lambda^2\left(\frac{S^2\ln S}{S^2-1}\right)-\lambda(1+\lambda)}$$

$$\mu = \frac{2S(S-1)-(S^2-1)}{2(1+2\lambda)S(S-1)-2\lambda S^2\ln S-(1+\lambda)(S^2-1)}$$

$$\mu = \frac{(S-1)^2}{2(1+2\lambda)S(S-1)-2\lambda S^2\ln S-(1+\lambda)(S^2-1)}$$

Finally, if $k \neq 0, 1, 2$,

$$\mu = \frac{\frac{k\lambda}{1-k}\left(\frac{S-S^k}{S^k-1}\right) - \lambda}{(1+2\lambda)\frac{k\lambda}{1-k}\left(\frac{S-S^k}{S^k-1}\right) - \frac{k\lambda^2}{2-k}\left(\frac{S^2-S^k}{S^k-1}\right) - \lambda(1+\lambda)}$$

$$\mu = \frac{\frac{k}{1-k}(S-S^k) - (S^k-1)}{(1+2\lambda)\frac{k}{1-k}(S-S^k) - \lambda\frac{k}{2-k}(S^2-S^k) - (1+2\lambda-\lambda)(S^k-1)}$$

$$\mu = \frac{\frac{k}{1-k}(S-S^k) - (S^k-1)}{(1+2\lambda)\left[\frac{k}{1-k}(S-S^k) - (S^k-1)\right] - \lambda\left[\frac{k}{2-k}(S^2-S^k) - (S^k-1)\right]}$$

$$\mu = \frac{\frac{k(S-1)-(S^k-1)}{1-k}}{(1+2\lambda)\left[\frac{k(S-1)-(S^k-1)}{1-k}\right] - \frac{\lambda}{2-k}\left[k(S^2-1) - 2(S^k-1)\right]}$$

$$\mu = \frac{k(S-1) - (S^k-1)}{(1+2\lambda)\left[k(S-1) - (S^k-1)\right] - \frac{\lambda(1-k)}{2-k}\left[2k(S-1) - 2(S^k-1) + k(S-1)^2\right]}$$

$$\mu = \frac{k(S-1) - (S^k-1)}{\left[(1+2\lambda) - \frac{2\lambda(1-k)}{2-k}\right]\left[k(S-1) - (S^k-1)\right] - \frac{\lambda(1-k)}{2-k}\left[k(S-1)^2\right]}$$

$$\mu = \frac{k(S-1) - (S^k-1)}{\left(1 + \frac{2\lambda}{2-k}\right)\left[k(S-1) - (S^k-1)\right] - \frac{\lambda k(1-k)}{2-k}(S-1)^2}$$

where we once again use Proposition 3.B.2. ■

Proposition 3.B.4 *For $\lambda > 0$,* $\overline{\mu} = \begin{cases} \frac{1}{\left(1+\frac{2\lambda}{2-k}\right) - \frac{k(1-k)/(2-k)}{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k - 1\right]}} & \text{, if } k \neq 0,1,2 \\ \frac{1}{(1+2\lambda) - \left[(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1\right]^{-1}} & \text{, if } k = 1 \\ \frac{1}{(1+2\lambda)(\lambda+1) - 2\lambda(\lambda+1)^2\ln\left(\frac{\lambda+1}{\lambda}\right)} & \text{, if } k = 2 \end{cases}$

Proof. Let $\lambda > 0$. Since $S \equiv \frac{\tilde{s}(\overline{A})}{\lambda} + 1$ and $\tilde{s}(A) \equiv 1 - \underline{A}/A_i$, $S = \frac{1-\tilde{A}^{-1}}{\lambda} + 1$. Hence, $S \to \frac{\lambda+1}{\lambda}$ when $\overline{A} \to +\infty$, implying $\overline{\mu} \equiv \lim_{\overline{A}\to+\infty} \mu = \lim_{S\to\frac{\lambda+1}{\lambda}} \mu$. As a result, if $k = 1$, from Proposition 3.B.3,

$$\lim_{\overline{A}\to+\infty} \mu = \lim_{S\to\frac{\lambda+1}{\lambda}} \frac{S\ln S - (S-1)}{(1+2\lambda)\left[S\ln S - (S-1)\right] - \lambda(S-1)^2}$$

$$\lim_{\overline{A}\to+\infty} \mu = \frac{(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1}{(1+2\lambda)\left[(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1\right] - 1} = \frac{1}{(1+2\lambda) - \frac{1}{(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1}}$$

If $k = 2$, using Proposition 3.B.3,

$$\lim_{\overline{A}\to+\infty} \mu = \lim_{S\to\frac{\lambda+1}{\lambda}} \frac{(S-1)^2}{2(1+2\lambda)S(S-1) - 2\lambda S^2 \ln S - (1+\lambda)(S^2-1)}$$

$$\lim_{\overline{A}\to+\infty} \mu = \frac{\frac{1}{\lambda^2}}{2(1+2\lambda)\left(\frac{\lambda+1}{\lambda}\right)\left(\frac{1}{\lambda}\right) - 2\lambda\left(\frac{\lambda+1}{\lambda}\right)^2 \ln\left(\frac{\lambda+1}{\lambda}\right) - (1+\lambda)\left(\frac{1+2\lambda}{\lambda}\right)\left(\frac{1}{\lambda}\right)}$$

$$\lim_{\overline{A}\to+\infty} \mu = \frac{1}{(1+2\lambda)(\lambda+1) - 2\lambda(\lambda+1)^2 \ln\left(\frac{\lambda+1}{\lambda}\right)}$$

Finally, if $k \neq 0, 1, 2$,

$$\lim_{\overline{A}\to+\infty} \mu = \lim_{S\to\frac{\lambda+1}{\lambda}} \frac{k(S-1) - (S^k - 1)}{\left(1 + \frac{2\lambda}{2-k}\right)[k(S-1) - (S^k-1)] - \frac{\lambda k(1-k)}{2-k}(S-1)^2}$$

$$\lim_{\overline{A}\to+\infty} \mu = \frac{1}{\left(1+\frac{2\lambda}{2-k}\right) - \frac{\left[\frac{\lambda k(1-k)}{2-k}\right]\frac{1}{\lambda^2}}{\frac{k}{\lambda} - \left[\left(\frac{\lambda+1}{\lambda}\right)^k - 1\right]}} = \frac{1}{\left(1+\frac{2\lambda}{2-k}\right) - \frac{k(1-k)/(2-k)}{k - \lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k - 1\right]}}$$

where we again use Proposition 3.B.3. ■

Proposition 3.B.5 *For $\lambda > 0$ and $k \in (0, 2]$, $\frac{\partial \overline{\mu}}{\partial \lambda} > 0$.*

Proof. Let $\lambda > 0$. It is sufficient to show that $\frac{\partial(1/\overline{\mu})}{\partial \lambda} < 0$ for $k \in (0, 2]$. If $k = 1$, from Proposition 3.B.4,

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{\partial\left[(1+2\lambda) - \left[(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1\right]^{-1}\right]}{\partial\lambda}$$

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = 2 + \frac{\ln\left(\frac{\lambda+1}{\lambda}\right) + \lambda\left(-\frac{1}{\lambda^2}\right)}{\left[(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1\right]^2} = \frac{2\left[\left(\frac{\lambda+1}{\lambda}\right)\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda}\right]^2 + \frac{1}{\lambda^2}\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda^3}}{\lambda^{-2}\left[(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1\right]^2}$$

Being $y \equiv \frac{\lambda+1}{\lambda} \to \lambda = \frac{1}{y-1}$,

$$f(y) \equiv 2\left[\left(\frac{\lambda+1}{\lambda}\right)\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda}\right]^2 + \frac{1}{\lambda^2}\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda^3}$$

$$f(y) = 2\left[y\ln y - (y-1)\right]^2 + (y-1)^2 \ln y - (y-1)^3 \to f(1) = 0$$

$$
\begin{aligned}
f'(y) =& 4\left[y\ln y-(y-1)\right]\ln y+2(y-1)\ln y+\frac{(y-1)^2}{y}-3(y-1)^2\\
f'(y) =& 2\left[2y\ln y-(y-1)\right]\ln y+(y-1)^2\left(\frac{1}{y}-3\right)\rightarrow f'(1)=0
\end{aligned}
$$

$$
\begin{aligned}
f''(y) =& 2\left[2\ln y-\left(\frac{y-1}{y}\right)\right]+2\left(2\ln y+1\right)\ln y+2(y-1)\left(\frac{1}{y}-3\right)\\
&+(y-1)^2\left(-\frac{1}{y^2}\right)\\
f''(y) =& \ln y\left(4\ln y+6\right)-\left(\frac{y-1}{y}\right)^2-6(y-1)\rightarrow f''(1)=0
\end{aligned}
$$

$$
\begin{aligned}
f'''(y) =& \left(\frac{4\ln y+6}{y}\right)+4\frac{\ln y}{y}-2\left(\frac{y-1}{y}\right)\left(\frac{1}{y^2}\right)-6\\
f'''(y) =& 2y^{-3}\left[4y^2\ln y-(y-1)+3y^2-3y^3\right]
\end{aligned}
$$

$$
\begin{aligned}
g(y) \equiv& 4y^2\ln y-(y-1)+3y^2-3y^3\rightarrow g(1)=0\\
g'(y) =& 8y\ln y+4y-1+6y-9y^2=8y\ln y+10y-9y^2-1\rightarrow g'(1)=0\\
g''(y) =& 8\ln y+8+10-18y=8\ln y-18y+18\rightarrow g''(1)=0\\
g'''(y) =& 8y^{-1}-18
\end{aligned}
$$

Since $\lambda>0\rightarrow y>1$, $g'''(y)<0\xrightarrow{g''(1)=0}g''(y)<0\xrightarrow{g'(1)=0}g'(y)<0\xrightarrow{g(1)=0}g(y)<0\rightarrow f'''(y)<0\xrightarrow{f''(1)=0}f''(y)<0\xrightarrow{f'(1)=0}f'(y)<0\xrightarrow{f(1)=0}f(y)<0$. Thus, $\frac{\partial(1/\overline{\mu})}{\partial\lambda}<0$ for $k=1$.

If $k=2$, using Proposition 3.B.4,

$$
\begin{aligned}
\frac{\partial(1/\overline{\mu})}{\partial\lambda} =& \frac{\partial\left[\left(1+2\lambda\right)\left(\lambda+1\right)-2\lambda\left(\lambda+1\right)^2\ln\left(\frac{\lambda+1}{\lambda}\right)\right]}{\partial\lambda}\\
\frac{\partial(1/\overline{\mu})}{\partial\lambda} =& (1+4\lambda+2)-2\left[(\lambda+1)^2+2\lambda(\lambda+1)\right]\ln\left(\frac{\lambda+1}{\lambda}\right)-2\lambda^2(\lambda+1)\left(-\lambda^{-2}\right)\\
\frac{\partial(1/\overline{\mu})}{\partial\lambda} =& (6\lambda+5)-2(\lambda+1)(3\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right)\\
\frac{\partial(1/\overline{\mu})}{\partial\lambda} =& \lambda^2\left[5\left(\frac{\lambda+1}{\lambda}\right)\left(\frac{1}{\lambda}\right)+\left(\frac{1}{\lambda}\right)-2\left(\frac{\lambda+1}{\lambda}\right)\left(\frac{\lambda+1}{\lambda}+2\right)\ln\left(\frac{\lambda+1}{\lambda}\right)\right]
\end{aligned}
$$

Again, let $y \equiv \frac{\lambda+1}{\lambda} \to \lambda = \frac{1}{y-1}$. However, in this case let $f(y) \equiv 5\left(\frac{\lambda+1}{\lambda}\right)\left(\frac{1}{\lambda}\right)+\left(\frac{1}{\lambda}\right)-2\left(\frac{\lambda+1}{\lambda}\right)\left(\frac{\lambda+1}{\lambda}+2\right)\ln\left(\frac{\lambda+1}{\lambda}\right) = 5y(y-1)+(y-1)-2y(y+2)\ln y \to f(1)=0$. As a result, $f'(y) = 10y-5+1-2(2y+2)\ln y-2(y+2) = 8(y-1)-4(y+1)\ln y \to f'(1) = 0$, $f''(y) = 8-4\ln y-4\left(\frac{y+1}{y}\right) \to f''(1)=0$, and $f'''(y) = -\frac{4}{y}+\frac{4}{y^2} = 4\left(\frac{1-y}{y^2}\right)$. Since $\lambda > 0 \to y > 1$, $f'''(y) < 0 \xrightarrow{f''(1)=0} f''(y) < 0 \xrightarrow{f'(1)=0} f'(y) < 0 \xrightarrow{f(1)=0} f(y) < 0$, implying $\frac{\partial(1/\overline{\mu})}{\partial\lambda} < 0$ for $k=2$.

If $k \neq 0,1,2$, using again Proposition 3.B.4,

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{\partial\left[\left(1+\frac{2\lambda}{2-k}\right)-\frac{k(1-k)/(2-k)}{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]}\right]}{\partial\lambda}$$

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{2}{2-k}+\left[\frac{k(1-k)}{2-k}\right]\frac{-\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]-\lambda\left[k\left(\frac{\lambda+1}{\lambda}\right)^{k-1}\left(-\frac{1}{\lambda^2}\right)\right]}{\left\{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}^2}$$

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{2}{2-k}+\left[\frac{k(1-k)}{2-k}\right]\frac{\frac{k}{\lambda}\left(\frac{\lambda+1}{\lambda}\right)^{k-1}-\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]}{\left\{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}^2}$$

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{\frac{2}{2-k}\left\{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}^2+\frac{k(1-k)}{2-k}\left\{\frac{k}{\lambda}\left(\frac{\lambda+1}{\lambda}\right)^{k-1}-\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}}{\left\{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}^2}$$

Consider now that $k \in (0,2)\setminus\{1\}$, when $2-k>0$. Let $y \equiv \frac{\lambda+1}{\lambda}$ again and

$$\begin{aligned} f(y) \equiv & 2\left\{k-\lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\}^2 \\ & -k(k-1)\left\{\frac{k}{\lambda}\left(\frac{\lambda+1}{\lambda}\right)^{k-1}-\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]\right\} \\ f(y) = & 2\left[k-\left(\frac{y^k-1}{y-1}\right)\right]^2-k(k-1)\left[k(y-1)y^{k-1}-\left(y^k-1\right)\right] \to \lim_{y\to 1^+} f(y)=0 \end{aligned}$$

$$\begin{aligned} f'(y) = & -4\left[k-\left(\frac{y^k-1}{y-1}\right)\right]\left[\frac{ky^{k-1}(y-1)-(y^k-1)}{(y-1)^2}\right] \\ & -k(k-1)\left[k^2y^{k-1}-k(k-1)y^{k-2}-ky^{k-1}\right] \end{aligned}$$

$$f'(y) = \frac{-4}{y-1}\left[k - \left(\frac{y^k-1}{y-1}\right)\right]\left[k - \left(\frac{y^k-1}{y-1}\right) + k(y^{k-1}-1)\right] - k^2(k-1)^2(y-1)y^{k-2}$$

$$f'(y) = \frac{4\left[k - \left(\frac{y^k-1}{y-1}\right)\right]^2 + 4\left[k - \left(\frac{y^k-1}{y-1}\right)\right]k(y^{k-1}-1) + k^2(k-1)^2(y-1)^2y^{k-2}}{-(y-1)}$$

$$f'(y) = \frac{\left[2\left(1 - \frac{y^k-1}{k(y-1)}\right) + (y^{k-1}-1)\right]^2 + \left[(k-1)^2(y-1)^2y^{k-2} - (y^{k-1}-1)^2\right]}{-k^{-2}(y-1)}$$

$$g(y) \equiv (k-1)^2(y-1)^2y^{k-2} - (y^{k-1}-1)^2 \rightarrow g(1) = 0$$
$$g'(y) = 2(k-1)^2(y-1)y^{k-2} + (k-1)^2(y-1)^2(k-2)y^{k-3} - 2(y^{k-1}-1)(k-1)y^{k-2}$$

$$h(y) \equiv \frac{g(y)}{y^{k-3}} = 2(k-1)^2(y^2-y) + (k-1)^2(k-2)(y-1)^2 - 2(k-1)(y^k-y)$$
$$h'(y) = 2(k-1)^2(2y-1) + 2(k-1)^2(k-2)(y-1) - 2(k-1)(ky^{k-1}-1)$$
$$h''(y) = 4(k-1)^2 + 2(k-1)^2(k-2) - 2k(k-1)^2y^{k-2}$$
$$\therefore h(1) = h'(1) = h''(1) = 0$$
$$h'''(y) = -2k(k-1)^2(k-2)y^{k-3} = 2(k-1)^2y^{k-3}k(2-k)$$

Since $\lambda > 0 \rightarrow y > 1$ and $k \in (0,2)\setminus\{1\} \rightarrow k(2-k) > 0$, $h'''(y) > 0 \xrightarrow{h''(1)=0} h''(y) > 0 \xrightarrow{h'(1)=0} h'(y) > 0 \xrightarrow{h(1)=0} h(y) > 0 \rightarrow g'(y) > 0 \xrightarrow{g(1)=0} g(y) > 0 \rightarrow f'(y) < 0 \xrightarrow{\lim_{y\to 1^+} f(y)=0} f(y) < 0$. Hence, $\frac{\partial(1/\overline{\mu})}{\partial\lambda} < 0$ for $k \in (0,2)\setminus\{1\}$.

In short, $\frac{\partial(1/\overline{\mu})}{\partial\lambda} < 0$ for $k \in (0,2]$. ■

Proposition 3.B.6 *For $k \neq 0$ and $\lambda > 0$, $\lim_{\lambda\to 0^+}\overline{\mu} = \max\{2-k, 1\}$.*

Proof. Let $\lambda > 0$. It is sufficient to show that $\lim_{\lambda\to 0^+}(1/\overline{\mu}) = (2-k)^{-1}$ if $k < 1$, $k \neq 0$, and $\lim_{\lambda\to 0^+}(1/\overline{\mu}) = 1$ if $k \geq 1$. If $k = 1$, from Proposition 3.B.4,

$$\lim_{\lambda\to 0^+}(1/\overline{\mu}) = 1 + \frac{1}{1 - \lim_{\lambda\to 0^+}\ln\left(\frac{\lambda+1}{\lambda}\right)} = 1$$

If $k = 2$, using Proposition 3.B.4,

$$\lim_{\lambda\to 0^+}(1/\overline{\mu}) = 1 - 2\lim_{\lambda\to 0^+}\left[\frac{\ln\left(\frac{\lambda+1}{\lambda}\right)}{\lambda^{-1}}\right] = 1 - 2\lim_{\lambda\to 0^+}\left[\frac{\left(\frac{\lambda}{\lambda+1}\right)(-\lambda^{-2})}{-\lambda^{-2}}\right] = 1 - 0 = 1$$

where we apply L'Hôpital's rule to get the second equality.

Lastly, if $k \neq 0, 1, 2$, using again Proposition 3.B.4,

$$\lim_{\lambda\to 0^+}(1/\overline{\mu}) = 1 - \frac{k(1-k)/(2-k)}{k - \lim_{\lambda\to 0^+}\left[\frac{\left(\frac{\lambda+1}{\lambda}\right)^k}{\lambda^{-1}}\right]}$$

$$\lim_{\lambda\to 0^+}(1/\overline{\mu}) = \begin{cases} 1 - \frac{k(1-k)}{k(2-k)} = \frac{1}{2-k} & \text{, if } k < 0 \\ 1 - \frac{k(1-k)/(2-k)}{k - \lim_{\lambda\to 0^+}\left[\frac{k\left(\frac{\lambda+1}{\lambda}\right)^{k-1}(-\lambda^{-2})}{-\lambda^{-2}}\right]} = 1 - \frac{k(1-k)}{k(2-k)} = \frac{1}{2-k} & \text{, if } k \in (0,1) \\ 1 - \frac{k(1-k)/(2-k)}{k - \lim_{\lambda\to 0^+}\left[\frac{k\left(\frac{\lambda+1}{\lambda}\right)^{k-1}(-\lambda^{-2})}{-\lambda^{-2}}\right]} = 1 - 0 = 1 & \text{, if } k > 1, k \neq 2 \end{cases}$$

where we use L'Hôpital's rule for $k \in (0, 1)$ and $k > 1$, $k \neq 2$. ■

Proposition 3.B.7 *For $k \neq 0$ and $\lambda > 0$, $\lim_{\lambda\to+\infty}\overline{\mu} = 3$.*

Proof. Let $\lambda > 0$. It is sufficient to show that $\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1/3$ for $k \neq 0$. If $k = 1$, from Proposition 3.B.4,

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = \lim_{\lambda\to+\infty}\left[1 + \frac{2\lambda(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 2\lambda - 1}{(\lambda+1)\ln\left(\frac{\lambda+1}{\lambda}\right) - 1}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \lim_{\lambda\to+\infty}\left[\frac{2\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{2\lambda+1}{\lambda(\lambda+1)}}{\frac{1}{\lambda}\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{1}{\lambda(\lambda+1)}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \lim_{\lambda\to+\infty}\left[\frac{2\left(\frac{\lambda}{\lambda+1}\right)\left(-\frac{1}{\lambda^2}\right) - \frac{2\lambda(\lambda+1)-(2\lambda+1)(2\lambda+1)}{\lambda^2(\lambda+1)^2}}{-\frac{1}{\lambda^2}\ln\left(\frac{\lambda+1}{\lambda}\right) + \frac{1}{\lambda}\left(\frac{\lambda}{\lambda+1}\right)\left(-\frac{1}{\lambda^2}\right) + \frac{2\lambda+1}{\lambda^2(\lambda+1)^2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \lim_{\lambda\to+\infty}\left[\frac{-\frac{2\lambda}{\lambda+1} - \frac{2\lambda}{\lambda+1} + \left(\frac{2\lambda+1}{\lambda+1}\right)^2}{-\ln\left(\frac{\lambda+1}{\lambda}\right) - \frac{\lambda+1}{(\lambda+1)^2} + \frac{2\lambda+1}{(\lambda+1)^2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =1+\lim_{\lambda\to+\infty}\left[\frac{\left(\frac{2\lambda+1}{\lambda+1}\right)^2-\frac{4\lambda}{\lambda+1}}{\frac{\lambda}{(\lambda+1)^2}-\ln\left(\frac{\lambda+1}{\lambda}\right)}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =1+\lim_{\lambda\to+\infty}\left[\frac{2\left(\frac{2\lambda+1}{\lambda+1}\right)\left(\frac{2(\lambda+1)-(2\lambda+1)}{(\lambda+1)^2}\right)-\frac{4(\lambda+1)-4\lambda}{(\lambda+1)^2}}{\frac{(\lambda+1)^2-2\lambda(\lambda+1)}{(\lambda+1)^4}-\left(\frac{\lambda}{\lambda+1}\right)\left(-\frac{1}{\lambda^2}\right)}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =1+\lim_{\lambda\to+\infty}\left[\frac{2\left(\frac{2\lambda+1}{\lambda+1}\right)-4}{\frac{\lambda+1}{\lambda}-\frac{\lambda-1}{\lambda+1}}\right]=1+\lim_{\lambda\to+\infty}\left[\frac{2\left(\frac{2(\lambda+1)-(2\lambda+1)}{(\lambda+1)^2}\right)}{-\frac{1}{\lambda^2}-\frac{(\lambda+1)-(\lambda-1)}{(\lambda+1)^2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =1+\lim_{\lambda\to+\infty}\left[\frac{\frac{2}{(\lambda+1)^2}}{-\frac{1}{\lambda^2}-\frac{2}{(\lambda+1)^2}}\right]=1-\lim_{\lambda\to+\infty}\left[\frac{2}{\left(\frac{\lambda+1}{\lambda}\right)^2+2}\right]=1-\frac{2}{3}=\frac{1}{3}$$

where we apply L'Hôpital's rule in the second, fourth, and fifth lines.

If $k=2$, using Proposition 3.B.4,

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[\frac{\frac{1+2\lambda}{\lambda(\lambda+1)}-2\ln\left(\frac{\lambda+1}{\lambda}\right)}{\lambda^{-1}(\lambda+1)^{-2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[\frac{\frac{2\lambda(\lambda+1)-(2\lambda+1)(2\lambda+1)}{\lambda^2(\lambda+1)^2}-2\left(\frac{\lambda}{\lambda+1}\right)\left(-\frac{1}{\lambda^2}\right)}{-\lambda^{-2}(\lambda+1)^{-2}-2\lambda^{-1}(\lambda+1)^{-3}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[\frac{2\lambda(\lambda+1)-(2\lambda+1)(\lambda+\lambda+1)+2\lambda(\lambda+1)}{-1-\frac{2\lambda}{\lambda+1}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[\frac{4\lambda(\lambda+1)-2\lambda^2-2\lambda(\lambda+1)-(2\lambda+1)}{-1-\frac{2\lambda}{\lambda+1}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[\frac{2\lambda(\lambda+1)-2\lambda^2-(2\lambda+1)}{-1-\frac{2\lambda}{\lambda+1}}\right]=\lim_{\lambda\to+\infty}\left[\frac{1}{1+\frac{2\lambda}{\lambda+1}}\right]=\frac{1}{3}$$

where we apply L'Hôpital's rule in the first line.

Finally, If $k\neq 0,1,2$, using again Proposition 3.B.4,

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =\lim_{\lambda\to+\infty}\left[1+\left(\frac{1}{2-k}\right)\frac{\frac{2k}{\lambda}-2\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]-\frac{k(1-k)}{\lambda^2}}{\frac{k}{\lambda^2}-\frac{1}{\lambda}\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) =1+\left(\frac{1}{2-k}\right)\lim_{\lambda\to+\infty}\left[\frac{-\frac{2k}{\lambda^2}-2k\left(\frac{\lambda+1}{\lambda}\right)^{k-1}\left(-\frac{1}{\lambda^2}\right)+2\frac{k(1-k)}{\lambda^3}}{-2\frac{k}{\lambda^3}+\frac{1}{\lambda^2}\left[\left(\frac{\lambda+1}{\lambda}\right)^k-1\right]-\frac{1}{\lambda}k\left(\frac{\lambda+1}{\lambda}\right)^{k-1}\left(-\frac{1}{\lambda^2}\right)}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \left(\frac{2k}{2-k}\right)\lim_{\lambda\to+\infty}\left[\frac{\left(\frac{\lambda+1}{\lambda}\right)^{k-1} - 1 + \frac{(1-k)}{\lambda}}{\left(\frac{\lambda+1}{\lambda}\right)^{k} - 1 + \frac{k}{\lambda}\left[\left(\frac{\lambda+1}{\lambda}\right)^{k-1} - 2\right]}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \left(\frac{2k}{2-k}\right)\lim_{\lambda\to+\infty}\left[\frac{-(k-1)\left(\frac{\lambda+1}{\lambda}\right)^{k-2} - (1-k)}{-2k\left[\left(\frac{\lambda+1}{\lambda}\right)^{k-1} - 1\right] - \frac{k}{\lambda}(k-1)\left(\frac{\lambda+1}{\lambda}\right)^{k-2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \left[\frac{2k(1-k)}{2-k}\right]\lim_{\lambda\to+\infty}\left[\frac{\left(\frac{\lambda+1}{\lambda}\right)^{k-2} - 1}{-2k\left[\left(\frac{\lambda+1}{\lambda}\right)^{k-1} - 1\right] - \frac{k}{\lambda}(k-1)\left(\frac{\lambda+1}{\lambda}\right)^{k-2}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \lim_{\lambda\to+\infty}\left[\frac{2k(1-k)\left(\frac{\lambda+1}{\lambda}\right)^{k-3}}{3k(k-1)\left(\frac{\lambda+1}{\lambda}\right)^{k-2} + \frac{k}{\lambda}(k-1)(k-2)\left(\frac{\lambda+1}{\lambda}\right)^{k-3}}\right]$$

$$\lim_{\lambda\to+\infty}(1/\overline{\mu}) = 1 + \frac{2k(1-k)}{3k(k-1)} = 1 - \frac{2}{3} = \frac{1}{3}$$

where we apply L'Hôpital's rule in the second, fourth, and sixth lines. ■

Proposition 3.B.8 *For $k \in (0,2]$ and $\lambda > 0$, $\tilde{A} \equiv \overline{A}/\underline{A}$, $\tilde{A} > 1$, is a continuous, strictly increasing, and well-defined function of μ if and only if $\mu \in (1,3)$ and $\lambda > \lambda^*(\mu)$, where $\lambda^*(\mu) \equiv \arg\min_{x\in[0,+\infty]} \lim_{\lambda\to x} |\overline{\mu} - \mu|$.*

Proof. Let $\hat{g}_s$ be the density of a truncated Lomax distribution with shape parameter $k \in (0,2]$ and scale parameter $\lambda > 0$ defined over $\tilde{s}(A) \in [0, \tilde{s}(A_h)]$, $A_h > \overline{A} > \underline{A}$, with $\hat{G}_s$ being the respective cumulative distribution function. It is easy to see $\tilde{g}_s(\tilde{s}(A)) = \hat{g}_s(\tilde{s}(A))/\hat{G}_s(\tilde{s}(\overline{A}))$ is the density of a truncated Lomax distribution with the same parameters $k \in (0,2]$ and $\lambda > 0$ over the support $[0, \tilde{s}(\overline{A})]$.[22] Given the density $\tilde{g}_s$ ($\hat{g}_s$), one gets the respective density for $A \in \left[\underline{A}, \overline{A}\right]$ ($A \in [\underline{A}, A_h]$), which we denote by $\tilde{g}$ ($\hat{g}$), with cumulative distribution function $\tilde{G}$ ($\hat{G}$). To get these distributions, note for any $A^* \in [\underline{A}, \overline{A}]$

$$P(A < A^*) = P(\underline{A}/A^* < \underline{A}/A) = P(1 - \underline{A}/A < 1 - \underline{A}/A^*) = P(\tilde{s}(A) < \tilde{s}(A^*))$$

[22] After all, the Lomax density is the density of a Pareto with the support shifted by $\lambda > 0$.

implying $\tilde{G}(A) = \tilde{G}_s(\tilde{s}(A)) \to \tilde{g}(A) = \tilde{g}_s(\tilde{s}(A))\underline{A}/A^2$ and $\hat{G}(A) = \hat{G}_s(\tilde{s}(A)) \to \hat{g}(A) = \hat{g}_s(\tilde{s}(A))\underline{A}/A^2$. Given these distributions and since $\tilde{g}_s(\tilde{s}(A)) = \frac{\hat{g}_s(\tilde{s}(A))}{\hat{G}_s(\tilde{s}(\overline{A}))}$,

$$\frac{\hat{g}(A)}{\hat{G}(\overline{A})} = \left[\frac{\hat{g}_s(\tilde{s}(A))}{\hat{G}_s(\tilde{s}(\overline{A}))}\right]\underline{A}/A^2 = \tilde{g}_s(\tilde{s}(A))\underline{A}/A^2 = \tilde{g}(A)$$

which allows us to use Proposition 3.B.1. Thus, $\overline{A}$, $\overline{A} > \underline{A}$ is continuous, strictly increasing, and well defined in μ if and only if $\mu \in (1, \overline{\mu})$, where $\overline{\mu}$ is given in Proposition 3.B.4. Since μ is only a function of $\tilde{A} \equiv \overline{A}/\underline{A}$ (given k and λ) due to Proposition 3.B.3 and the fact that $S \equiv \frac{\tilde{s}(\overline{A})}{\lambda} + 1 = \frac{1-\tilde{A}^{-1}}{\lambda} + 1$, these features of $\overline{A}$ also hold for $\tilde{A}$.

From Propositions 3.B.5, 3.B.6, and 3.B.7, the image of $\overline{\mu}$ over $\lambda \in (0, +\infty)$ is $(\max\{2-k, 1\}, 3)$ for $k \in (0, 2]$. Hence, any $k \in (0, 2]$ is viable in the sense that there is always $\lambda > 0$ such as $\mu \in (1, \overline{\mu})$ if and only if $\mu \in (1, 3)$. Under such condition, from Proposition 3.B.5, any $\lambda > \lambda^*(\mu)$ is viable, where $\lambda^*(\mu) \equiv \arg\min_{x\in[0,+\infty]} \lim_{\lambda\to x} |\overline{\mu} - \mu|$ for $k \in (0, 2]$, which is well defined as $\overline{\mu}$ is a continuous and monotonic function of λ (Proposition 3.B.5).[23] Therefore, for $k \in (0, 2]$ and $\lambda > 0$, $\mu \in (1, \overline{\mu})$ holds if and only if $\mu \in (1, 3)$ and $\lambda > \lambda^*(\mu)$. ■

Figure 3.B.1 illustrates the results of Proposition 3.B.8, plotting $\tilde{A}$ against μ for truncated Lomax distributions with $k = 1$ and $\lambda = 10, 1, 0.1, 0.01$.

We finish this section with a comment about the generalization of Propositions 3.B.5 and 3.B.8 for any $k \neq 0$. In Proposition 3.B.5, we show

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{2}{2-k} + \left[\frac{k(1-k)}{2-k}\right] \frac{\frac{k}{\lambda}\left(\frac{\lambda+1}{\lambda}\right)^{k-1} - \left[\left(\frac{\lambda+1}{\lambda}\right)^k - 1\right]}{\left\{k - \lambda\left[\left(\frac{\lambda+1}{\lambda}\right)^k - 1\right]\right\}^2}$$

for any $k \neq 0, 1, 2$. Consequently, if $k = -1$,

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{2}{3}\left\{1 - \frac{-\frac{1}{\lambda}\left(\frac{\lambda}{\lambda+1}\right)^2 - \left(\frac{\lambda}{\lambda+1}\right) + 1}{\left[-1 - \lambda\left(\frac{\lambda}{\lambda+1}\right) + \lambda\right]^2}\right\} = \frac{2}{3}\left\{1 - \frac{\frac{1}{\lambda+1}\left[(\lambda+1) - \lambda - \left(\frac{\lambda}{\lambda+1}\right)\right]}{\left[(\lambda-1) - \frac{\lambda^2}{\lambda+1}\right]^2}\right\}$$

[23] Note $\overline{\mu} = \mu$ under $\lambda = \lambda^*(\mu) \in (0, +\infty)$ except for $k \in (0, 1)$ and $\mu \in (1, 2-k]$, when $\lambda^*(\mu) = 0$ as $\overline{\mu}$ is strictly increasing in $\lambda > 0$ with $\overline{\mu} \in (2-k, 3)$.

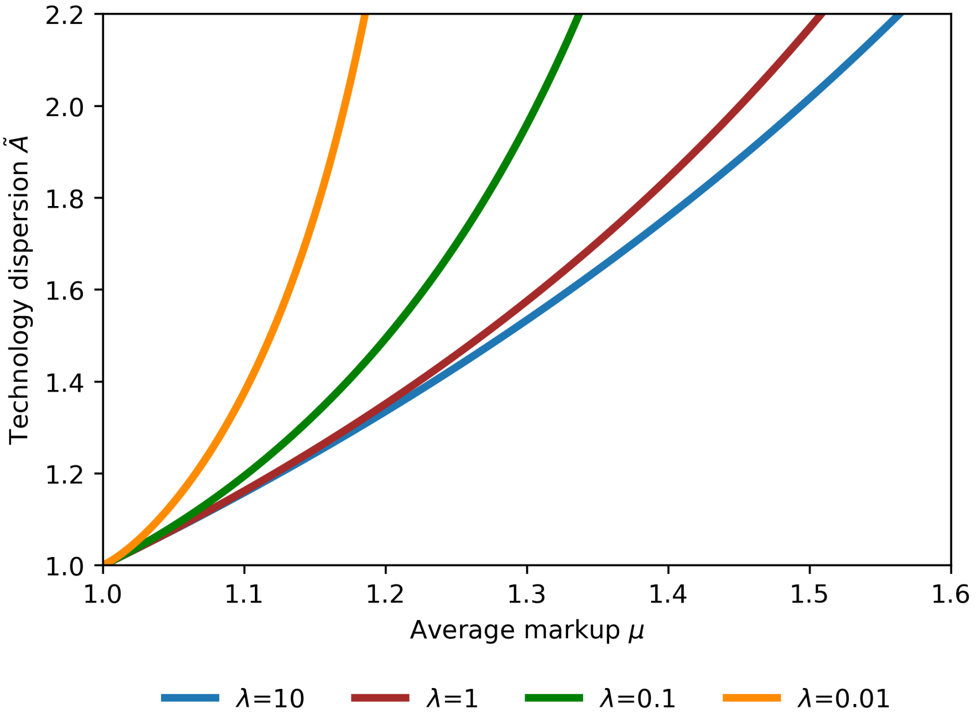


Figure 3.B.1 – Technology dispersion vs. average markup - Zipf's law calibration, $k = 1$.

$$\frac{\partial(1/\overline{\mu})}{\partial\lambda} = \frac{2}{3}\left\{1 - \frac{\frac{1}{\lambda+1}\left[\frac{(\lambda+1)-\lambda}{\lambda+1}\right]}{\left[\frac{(\lambda-1)(\lambda+1)-\lambda^2}{\lambda+1}\right]^2}\right\} = \frac{2}{3}\left\{1 - \frac{\frac{1}{\lambda+1}\left(\frac{1}{\lambda+1}\right)}{\left(\frac{-1}{\lambda+1}\right)^2}\right\} = \frac{2}{3}(1-1) = 0$$

Moreover, numerical results using this expression, such as those shown in Figure 3.B.2, suggest $\frac{\partial(1/\overline{\mu})}{\partial\lambda} > 0$ for $k < -1$ and $\frac{\partial(1/\overline{\mu})}{\partial\lambda} < 0$ for $k \in (-1, 0) \cup (2, +\infty)$. Assuming these results indeed hold, one can generalize Proposition 3.B.5 for any $k \neq 0$: for $\lambda > 0$, $\frac{\partial\overline{\mu}}{\partial\lambda} > 0$ if $k > -1, k \neq 0$, $\frac{\partial\overline{\mu}}{\partial\lambda} = 0$ if $k = -1$, and $\frac{\partial\overline{\mu}}{\partial\lambda} < 0$ if $k < -1$.

Given this (possible) new Proposition 3.B.5, one can also generalize Proposition 3.B.8: for $k \neq 0$ and $\lambda > 0$, $\tilde{A} \equiv \overline{A}/\underline{A}$, $\tilde{A} > 1$, is continuous, strictly increasing, and well defined in μ if and only if (i) $\mu \in (1, 3)$, (ii) $\mu = 3$ and $k \leq -1$, or (iii)

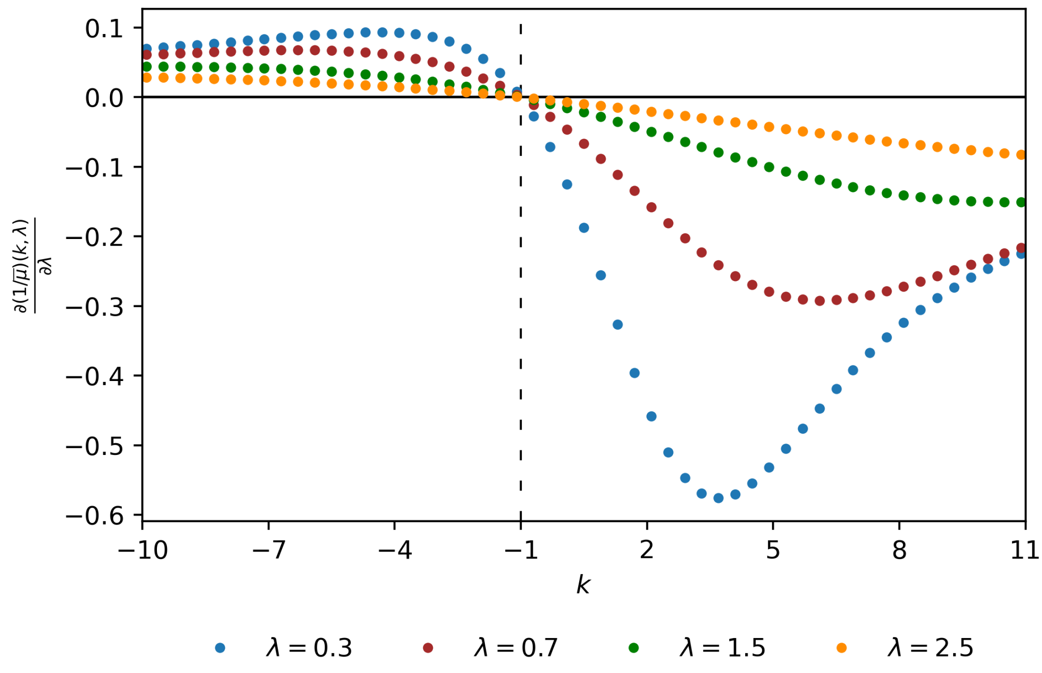


Figure 3.B.2 – $\frac{\partial(1/\overline{\mu})}{\partial\lambda}$ vs. k, $k \neq 0, 1, 2$.

$\mu > 3$ and $k < 2 - \mu < -1$, with any viable λ, that is, with

$$\lambda \in \begin{cases} (0, \lambda^*(\mu)) & \text{, if } k < -1 \\ (0, +\infty) & \text{, if } k = -1 \\ (\lambda^*(\mu), +\infty) & \text{, if } k > -1, k \neq 0 \end{cases}$$

where $\lambda^*(\mu) \equiv \arg\min_{x\in[0,+\infty]} \lim_{\lambda\to x} |\overline{\mu} - \mu|$ is now defined for any $k \notin \{-1, 0\}$.[24] After all, in this case, from this (possible) new Proposition 3.B.5 and Propositions 3.B.6 and 3.B.7, the image of $\overline{\mu}$ over $\lambda \in (0, +\infty)$ is

$$\begin{cases} (3, 2-k) & \text{, if } k < -1 \\ [3, 3] & \text{, if } k = -1 \\ (\max\{2-k, 1\}, 3) & \text{, if } k > -1, k \neq 0 \end{cases}$$

[24] Here, given this (possible) new Proposition 3.B.5, $\overline{\mu} = \mu$ under $\lambda^*(\mu) \in (0, +\infty)$ except in two cases. The first exception would be $k < -1$ and $\mu \in (1, 3]$, when $\lambda^*(\mu) = +\infty$ as $\overline{\mu}$ is strictly decreasing in $\lambda > 0$ with $\overline{\mu} \in (3, 2-k)$. The second exception would be $k \in (-1, 1) \setminus \{0\}$ and $\mu \in (1, 2-k]$, when $\lambda^* = 0$ as $\overline{\mu}$ is strictly increasing in $\lambda > 0$ with $\overline{\mu} \in (2-k, 3)$.

implying any $k \neq 0$ is viable if $\mu \in (1,3)$; if $\mu = 3$, $k \leq -1$; if $\mu > 3$, $\mu < 2 - k \rightarrow k < 2 - \mu < -1$. And, given a viable k, from this (possible) generalization of Proposition 3.B.5, one easily gets the viable λs shown in the statement of the (possible) new Proposition 3.B.8.

3.C Zipf's law calibration

3.C.1 Distributional assumption and the Zipf's law

Suppose $\tilde{s}(A) \equiv 1 - \underline{A}/A = \frac{s(A)}{\eta/q} \in \left[0, \tilde{s}(\overline{A})\right] = \left[0, 1 - \tilde{A}^{-1}\right]$ is truncated Lomax distributed with shape parameter $k \neq 0$ and scale parameter $\lambda > 0$ or, equivalently, $(\tilde{s}(A) + \lambda) \in [\lambda, \tilde{s}(\overline{A}) + \lambda]$ is truncated Pareto distributed with shape parameter $k \neq 0$. Thus, the density of $\tilde{s}(A)$ is $\tilde{g}_s(\tilde{s}(A)) = k\left[\frac{\lambda^k(\tilde{s}(\overline{A})+\lambda)^k}{(\tilde{s}(\overline{A})+\lambda)^k - \lambda^k}\right](\tilde{s}(A) + \lambda)^{-k-1}$, which is shown in Figure 3.C.1 for $k = 1$ and $\lambda = 0.2, 0.4, 1$, with $\tilde{s}(\overline{A}) = 1/3$.[25]

For $k \approx 1$ and $\lambda \approx 0$, firm market share is approximately truncated Pareto distributed with shape parameter $k \approx 1$, consistent with Zipf's law, particularly away from support's bounds. After all,

$$P\left(\tilde{s}(A) \geq x\right) = \int_x^{\tilde{s}(\overline{A})} \tilde{g}_s(y)dy = \frac{(x+\lambda)^{-k} - (\tilde{s}(\overline{A}) + \lambda)^{-k}}{\lambda^{-k} - (\tilde{s}(\overline{A}) + \lambda)^{-k}} \tag{3.C.1}$$

$$\frac{\partial \ln P\left(\tilde{s}(A) \geq x\right)}{\partial \ln x} = \frac{-k(x+\lambda)^{-k-1}x}{(x+\lambda)^{-k} - (\tilde{s}(\overline{A}) + \lambda)^{-k}} = -k\left[\frac{\frac{x}{x+\lambda}}{1 - \left(\frac{x+\lambda}{\tilde{s}(\overline{A})+\lambda}\right)^k}\right] \tag{3.C.2}$$

where we use in the last line $\frac{\partial x}{\partial \ln x} = \frac{\partial e^{\ln x}}{\partial \ln x} = e^{\ln x} = x$.[26] As a result, (i) $\frac{\partial \ln P(\tilde{s}(A) \geq x)}{\partial \ln x} \rightarrow 0$ when $x \rightarrow 0^+$, (ii) $\frac{\partial \ln P(\tilde{s}(A) \geq x)}{\partial \ln x} \rightarrow -\infty$ when $x \rightarrow \tilde{s}(\overline{A})^-$, and (iii) $\frac{\partial \ln P(\tilde{s}(A) \geq x)}{\partial \ln x} \approx$

[25] Alternatively, we could have imposed the distribution of $\theta_i = L_i/L = K_i/K$, when firm size would be measured by number of employees (or capital stock) instead of sales. Since we need $\mathrm{E}_a\left[(\underline{A}/A)^j\right]$ for $j \in \mathbb{N} \setminus \{0\}$ to compute the target moments, we should recover $\underline{A}/A_i$ from θ_i. Given Equation (3.2) for the continuous model and (3.A.1), $\theta_i = \frac{\overline{A}\Omega}{\underline{A}}(\eta/q)\left[(\underline{A}/A_i) - (\underline{A}/A_i)^2\right]$ and thus θ_i uniquely identifies $\underline{A}/A_i$ if and only if $\tilde{A} \equiv \overline{A}/\underline{A} \leq 2$. Thus, under this alternative strategy, one must impose $\tilde{A} \leq 2$, which may not hold empirically.

[26] From (3.C.1), $P\left(s(A) \geq x\right) = P\left(\tilde{s}(A) \geq \frac{x}{\eta/q}\right) = \frac{(x+\tilde{\lambda})^{-k} - (s(\overline{A})+\tilde{\lambda})^{-k}}{\tilde{\lambda}^{-k} - (s(\overline{A})+\tilde{\lambda})^{-k}}$ for $\tilde{\lambda} \equiv \lambda\,(\eta/q)$, meaning $s(A) \in [0, s(\overline{A})]$ is truncated Lomax distributed with shape parameter $k \neq 0$ and scale parameter $\tilde{\lambda} > 0$.

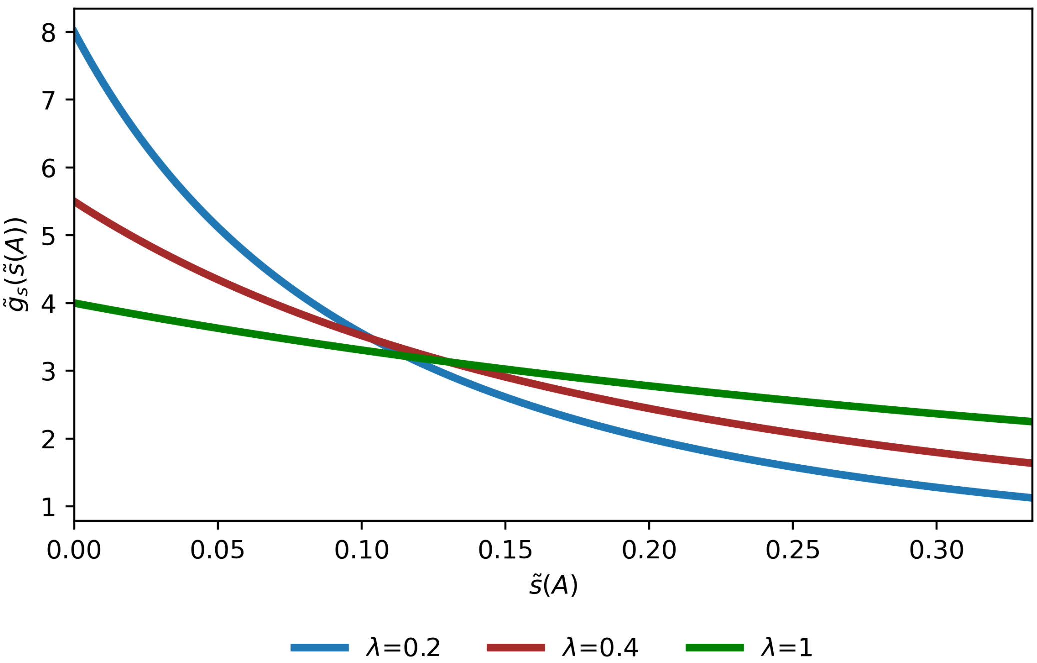


Figure 3.C.1 – Truncated Lomax density for $k = 1$ over the support $[0, 1/3]$.

$-k\left(\frac{x}{x+\lambda}\right)$ near the beginning of the support, when $\frac{x+\lambda}{\tilde{s}(\overline{A})+\lambda} \approx 0$. As a consequence, if $k \approx 1$ and $\lambda \approx 0$, $\frac{\partial \ln P(\tilde{s}(A) \geq x)}{\partial \ln x} \approx -1$ for x near the beginning of the support, as in Zipf's law.

It is worth mentioning that these deviations from Zipf's law near support bounds are not a shortcoming of the Lomax distribution. It would occur for any continuous truncated distribution. On the one hand, there is no distribution satisfying Zipf's law at $x = 0$ since $P(s \geq x) = cx^{-k}$ with $k \approx 1$ is only well defined for $x \neq 0$, implying $\frac{\partial \ln P(\tilde{s}(A) \geq x)}{\partial \ln x}$ should deviate from -1 sufficiently close to $x = 0$ as $P\left(\tilde{s}(A) \geq x\right)$ is a continuous function. On the other hand, to evaluate the behavior near the end of the support, assume, by contradiction, $P(s \geq x) = cx^{-k}$ for $s \in [\underline{s}, \overline{s}] \in (0, +\infty)$ and $k \approx 1$, with $c > 0$ as $P(s \geq x) > 0$ at least for $x = \underline{s}$. In this case, $P(s \geq \overline{s}) = c\overline{s}^{-k} > 0$, which is absurd as $s \in [\underline{s}, \overline{s}] \in (0, +\infty)$ and thus $P(s \geq \overline{s}) = 0$. Therefore, no truncated continuous distribution satisfies Zipf's law over a strictly positive support. A corollary of this result is that any distribution should deviate from Zipf's law near the end of its positive support. After all, if that were not true, by truncating this distribution from below, we would obtain a truncated continuous distribution over a narrow strictly positive support that satisfies Zipf's law, contradicting the previous result.

Under this distributional assumption, we show in Proposition 3.B.2 that

$$\mathrm{E}_a\left[(\tilde{s}(A)+\lambda)^j\right] = \begin{cases} \frac{k\lambda^j}{j-k}\left(\frac{S^j-S^k}{S^k-1}\right) & \text{, if } j \neq k \\ k\lambda^k\left(\frac{S^k \ln S}{S^k-1}\right) & \text{, if } j = k \end{cases} \tag{3.C.3}$$

for $k \neq 0$, $\lambda > 0$, and $j \in \mathbb{N} \setminus \{0\}$, where $S \equiv \frac{\tilde{s}(\overline{A})}{\lambda} + 1$.

3.C.2 Two-stage calibration procedure

We follow essentially the same two stages presented in Section 3.3.3 to quantify the model, but using (3.C.3) and the fact that $\tilde{s}(A) \equiv 1 - \underline{A}/A \rightarrow \mathrm{E}_a\left((\underline{A}/A)^j\right) = \mathrm{E}_a\left((1-\tilde{s}(A))^j\right)$ instead of Equation (3.9). In the first stage, we find $\underline{A}$ and $\overline{A}$ by matching aggregate TFP $\overline{A}\Omega$ and average markup μ to data. From Proposition 3.B.8, given $k \in (0,2]$, a unique solution to this problem exists if and only if $\mu \in (1,3)$ and $\lambda > \lambda^*(\mu)$, where $\lambda^*(\mu) \equiv \arg\min_{x\in[0,+\infty]} \lim_{\lambda\rightarrow x} |\overline{\mu} - \mu|$ and $\overline{\mu} \equiv \lim_{\overline{A}\rightarrow+\infty} \mu$ is computed from (3.10) in Proposition 3.B.4. Analogously to the baseline Pareto case, for a given distributional shape, allocative efficiency Ω is only a function of μ. Figure 3.C.2 plots this function Ω of μ for truncated Lomax distributions with $k = 1$ and $\lambda = 10, 1, 0.1, 0.01$, showing several noteworthy results. First, as before, Ω is strictly decreasing in $\mu = \frac{1-\alpha}{LS}$, with $\Omega \rightarrow 1^-$ when $\mu \rightarrow 1^+$. Second, given a time series of μ, a higher λ would imply a higher and less volatile estimated Ω.

In the second stage, we set $k \in (0,2]$, $k \approx 1$, which is viable as $\mu \in (1,3)$ in Brazil, and search a time-invariant λ that minimizes the distance to normalized $N_a HHI_L$.[27] We limit the search to $\lambda > \lambda^*(\max\{\mu\})$, where $\max\{\mu\}$ is the maximum μ in its time series.[28] Analogously to the baseline Pareto case, it is easy to see

[27] We also try to seek $k \neq 0$ and $\lambda > 0$ jointly in the second stage. Since, in this case, $k \in (0,2]$ may not hold, we rely on a (possible) generalization of Proposition 3.B.8 to any $k \neq 0$ discussed at the end of Appendix 3.B.2.

[28] Given $\mu \in (1,3)$ and $k \in (0,2]$, there is a solution for the first-stage algorithm if and only if $\lambda > \lambda^*(\mu)$. As a result, $\lambda > \max\{\lambda^*(\mu)\}$ ensure there is a solution for this algorithm in each

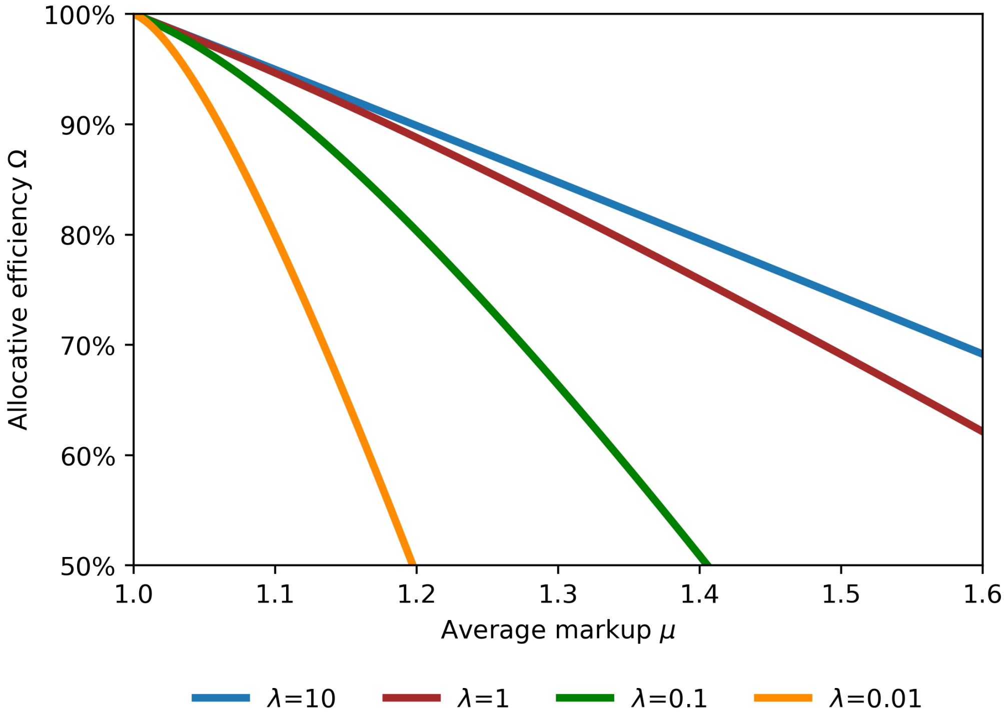


Figure 3.C.2 – Allocative efficiency vs. average markup - Zipf's law calibration, $k = 1$.

that TFP data are not required to estimate Ω, not even through λ. They are used only to pin down $\underline{A}$ and $\overline{A}$. Therefore, once more, the residual of the production function is not the TFP $\overline{A}\Omega$ itself, but rather only its technology component $\overline{A}$.

period, with $\max\{\lambda^*(\mu)\}$ being the maximum $\lambda^*(\mu)$ in its time series. However, since $\overline{\mu}$ is strictly increasing in λ for $k \in (0, 2]$ (Proposition 3.B.5), $\max\{\lambda^*(\mu)\} = \lambda^*(\max\{\mu\})$.

4 Zipf's law in the distribution of Brazilian firm size

4.1 Introduction

A power or scaling law holds for variables X and Y if $Y = cX^k$, where k is known as the power law exponent and c is typically an unremarkable constant. As Gabaix (2009), Gabaix (2016) points out, these power laws emerge in different domains, from natural phenomena (e.g., earthquakes, forest fires, and rivers), biology (e.g., Kleiber's law), and popularity of websites to economics, both in theory (e.g., the quantity theory of money) and empirically (e.g., Kaldor's stylized facts on economic growth). A power law may also apply to a distribution, with

$$P(S \geq s) = (\underline{s}/s)^k \tag{4.1}$$

for a random variable S, $S \geq \underline{s} > 0$, where $k > 0$. Generally, this distribution is known as Pareto (type I), but it is called Zipf's law when $k \approx 1$. In such cases, the probability of S being greater or equal to s is roughly proportional to $1/s$. This "law" was named after the linguist George Kingsley Zipf, who found analogous empirical regularity for the usage frequency of words in different languages and countries (Zipf, 1949), but it shows up in several other contexts. One illustrative example is the distribution of city size by population, especially among larger cities (Gabaix, 1999; Gabaix; Ioannides, 2004).[1]

In this paper, we evaluate Zipf's law for the distribution of firm size by the number of employees in Brazil. We use publicly available binned annual data from the Central Register of Enterprises (CEMPRE), which is held by the Brazilian Institute of Geography and Statistics (IBGE) and covers all formal organizations.

[1] For analogous evidence for Brazil, see Jr and Ribeiro (2006) and Justo (2014). On a related matter, Comitti, Shikida and Figueiredo (2022) estimate daily power law exponent k for the distribution of Brazilian municipalities by the number of infected people by COVID-19. Interestingly, they find it converges over time to 0.87, which is exactly the k they estimate for the distribution of municipality size by population.

Following the methodology proposed by Virkar and Clauset (2014), we find Zipf's law provides a very good, although not perfect, approximation to data for each year between 1996 and 2020 at the economy-wide level and also for agriculture, industry, and services alone. However, a lognormal distribution also performs well and even outperforms Zipf's law in certain cases.

Related literature. Empirical evidence supporting Zipf's law for firm size distribution has been found for several different countries with firm size measured by the number of employees, sales, income, total assets, and equity plus debt (Okuyama; Takayasu; Takayasu, 1999; Axtell, 2001; Fujiwara et al., 2004; Luttmer, 2007; Gabaix; Landier, 2008; Giovanni; Levchenko; Ranciere, 2011; Giovanni; Levchenko, 2013).[2] In particular, Giovanni and Levchenko (2013) use the ORBIS database to evaluate firm size distribution by total sales for a sample of 44 countries. In Giovanni and Levchenko (2013, p.295) own words,

> [...] the country sample is diverse: it includes major European economies (France, Germany, Netherlands), smaller E.U. accession countries (Czech Republic, Estonia), major middle income countries (Brazil, Argentina), as well as the two largest emerging markets (India and China). All in all, in this sample of 44 countries with very different characteristics, the distributions of firm size are remarkably consistent with Zipf's Law.

Specifically to Brazil, Silva et al. (2018) studies the distribution of firm size by net revenue, finding support for Zipf's law among the 1,000 largest firms in 2015.

The literature also shows contradictory evidence. For instance, there is some support for lognormality for firm size distribution (Stanley et al., 1995; Kondo; Lewis; Stella, 2023). Moreover, applying Lagrange multiplier tests, Resende and Cardoso (2022) find support to the more general Pareto type II and Pareto type IV against the Pareto type I and Zipf's law for firm size distribution by net revenue in Brazil.

[2] Fujiwara (2004) finds Zipf's law also holds for the distribution of total liabilities of *bankrupted* firms in Japan. For a survey of the empirical findings about Zipf's law for firm size, see Section 3 of Bottazzi, Pirino and Tamagni (2015).

The remainder of the paper proceeds as follows. Section 4.2 presents the data and methodology. Section 4.3 presents the empirical results. Finally, Section 4.4 concludes.

4.2 Data and methodology

We use publicly available annual data from CEMPRE, which is held by the IBGE and covers all formal organizations (corporate entities, public administration, and non-profit organizations). We split the analysis into two distinct periods due to a methodological break in the database, which (i) altered the criteria for identifying active firms and (ii) updated the industry classification. First, between 1996 and 2006, when the industries are classified according to the National Classification of Economic Activities (CNAE), a Brazilian classification derived from the ISIC Rev.3. Second, from 2006 to 2020, using CNAE 2.0, which follows the ISIC Rev.4. For both periods, we have the number of firms across all industries (up to 3-digit level) by nine size bins based on the number of employees: 0 to 4, 5 to 9, 10 to 19, 20 to 29, 30 to 49, 50 to 99, 100 to 249, 250 to 499, and 500 or more. All our analyses are done at the economy-wide level and also for agriculture, industry, and services alone.[3] Table 4.1 presents the data for these industries in selected years, showing the new criteria for identifying active firms substantially lowered the number of firms in 2006, notably for firms with up to four employees.

Virkar and Clauset (2014) suggest three steps to evaluate the prevalence of a distributional power law in binned data: (i) fit the power law, (ii) test the power law's plausibility, and (iii) compare against alternative distributions.[4] We follow similar steps. Our alternative distributions are (i) a strong Zipf's law or simply a Zipf distribution, that is, a Pareto density with $k = 1$, and (ii) a lognormal density. The choice of the lognormal is due to two reasons. First, the "[..] lognormal provides a strong test because for a wide range of sample sizes it produces bin counts that are reasonably power-law-like when plotted on log–log axes [...]" (Virkar; Clauset, 2014, p.103). Second, there is also evidence supporting lognormality for

[3] For the CNAE, we classified sections A and B as agriculture, C to F as industry, and G to Q as services. For the CNAE 2.0, A is agriculture, B to F is industry, and G to U is services.

[4] For an analogous approach for non-binned data, see Clauset, Shalizi and Newman (2009).

Table 4.1 – Number of firms by firm size

Number of employees	1996-2006 database 1996	1996-2006 database 2001	1996-2006 database 2006	2006-2020 database 2006	2006-2020 database 2013	2006-2020 database 2020
All industries						
0 to 4	2,616,788	3,903,486	4,730,580	3,324,519	3,985,367	4,090,186
5 to 9	327,372	432,626	542,426	531,612	755,609	739,242
10 to 19	141,337	193,133	265,581	261,271	379,902	358,736
20 to 29	40,693	55,032	69,486	69,433	102,152	93,372
30 to 49	31,260	39,498	50,276	50,222	73,368	65,053
50 to 99	23,133	27,102	33,294	33,269	47,651	43,294
100 to 249	15,244	16,732	19,683	19,664	27,132	24,341
250 to 499	5,713	6,283	7,807	7,801	10,429	9,739
500 or more	5,181	5,933	7,793	7,787	10,624	10,128
Total	3,206,721	4,679,825	5,726,926	4,305,578	5,392,234	5,434,091
Agriculture						
0 to 4	16,419	23,666	38,961	21,850	93,237	89,402
5 to 9	3,436	3,737	4,681	4,249	5,870	6,638
10 to 19	1,909	2,160	2,948	2,740	3,686	3,657
20 to 29	735	814	980	977	1,105	1,136
30 to 49	583	717	778	760	863	811
50 to 99	447	538	585	599	637	681
100 to 249	247	310	404	402	398	407
250 to 499	103	132	121	125	157	160
500 or more	88	124	127	127	127	130
Total	23,967	32,198	49,585	31,829	106,080	103,022
Industry						
0 to 4	315,907	413,192	474,964	314,128	433,166	472,907
5 to 9	61,262	73,224	83,092	82,158	118,577	107,997
10 to 19	36,803	48,727	59,429	59,166	79,931	69,686
20 to 29	13,656	18,474	21,407	21,664	29,726	24,498
30 to 49	11,487	14,795	17,571	17,588	23,142	17,941
50 to 99	9,045	10,906	13,200	13,231	17,366	13,127
100 to 249	5,759	6,160	7,308	7,295	9,836	7,399
250 to 499	2,089	1,942	2,438	2,423	3,228	2,596
500 or more	1,726	1,622	2,034	2,038	2,902	2,283
Total	457,734	589,042	681,443	519,691	717,874	718,434
Services						
0 to 4	2,284,462	3,466,628	4,216,655	2,988,541	3,458,964	3,527,877
5 to 9	262,674	355,665	454,653	445,205	631,162	624,607
10 to 19	102,625	142,246	203,204	199,365	296,285	285,393
20 to 29	26,302	35,744	47,099	46,792	71,321	67,738
30 to 49	19,190	23,986	31,927	31,874	49,363	46,301
50 to 99	13,641	15,658	19,509	19,439	29,648	29,486
100 to 249	9,238	10,262	11,971	11,967	16,898	16,535
250 to 499	3,521	4,209	5,248	5,253	7,044	6,983
500 or more	3,367	4,187	5,632	5,622	7,595	7,715
Total	2,725,020	4,058,585	4,995,898	3,754,058	4,568,280	4,612,635

Source: publicly available CEMPRE database.

firm size distribution (Stanley et al., 1995; Kondo; Lewis; Stella, 2023). Given these alternative distributions, we consider the following three steps to evaluate power and strong Zipf's law:

1. Fit Pareto and lognormal distributions.

2. Test Pareto, lognormal, and strong Zipf's law plausibility.

3. Compare Pareto, Zipf, and lognormal distributions.

In the following, we present the methodology used in each of these three steps.

4.2.1 Step 1: fitting the distributions

Before discussing the estimators, three comments are in order. First, in some empirical applications, a distributional power law may hold but only in the upper tail, meaning one must also gauge the support's lower bound $\underline{s} > 0$. For instance, one can visually identify the point beyond which the empirical survival function becomes roughly straight on a log-log plot, although more objective methods also exist (Clauset; Shalizi; Newman, 2009; Virkar; Clauset, 2014; Schluter, 2021). However, since we have just a few bins, we choose to test all possible $\underline{s}$ instead of choosing a specific one, setting $\underline{s} = 5, 10, 20, 30, 50$ for both Pareto and lognormal distributions.[5] Second, since lognormal's support begins at zero and we need it to start at $\underline{s} > 0$, we shift its density to the right by $\underline{s}$, supposing $S - \underline{s} > 0$ is lognormally distributed. Third, our measure of firm size, the number of employees, is discrete, whereas both Pareto and lognormal distributions have continuous supports. We address this issue by discretizing each distribution, defining the probability mass function as $P(S = s) \equiv P(S \geq s) - P(S \geq s + 1)$ for $s \in \{s \in \mathbb{N} | s \geq \underline{s}\}$, where $P(S \geq s)$ is computed from the respective continuous distribution.[6] This discretization is adopted by Kondo, Lewis and Stella (2023) and advocated, for the Pareto case, by Buddana and Kozubowski (2014). Differently, Clauset, Shalizi and

[5] We do not set $\underline{s} = 0$ because the Pareto support is strictly positive, while $\underline{s} = 100, 250, 500$ are discarded as we need at least four bins to ensure some degree of freedom in the lognormal estimation.

[6] Consequently, the probabilities add up to one by construction as $\sum_{s=\underline{s}}^{\infty} P(S = s) = P(S \geq \underline{s}) = 1$.

Newman (2009) consider a power law for the probability mass function assuming $P(S = s) \equiv \zeta(k, \underline{s})S^{-k-1}$, where ζ is a generalized zeta function, "which is rather inconvenient to work within an applied setting" (Buddana; Kozubowski, 2014, p.144).

We use two estimators for the Pareto distribution. First, we apply Ordinary Least Squares (OLS) to Equation (4.1), when we replace the survival function $P(S \geq s)$ by its empirical counterpart $\hat{P}(S \geq s)$, computed as the ratio between the number of firms with size $S \geq s$ and the number of firms with size $S \geq \underline{s}$.[7] Formally, given $\exp(\epsilon) \equiv \frac{\hat{P}(S\geq s)}{P(S\geq s)}$, the regression equation is

$$\ln \hat{P}(S \geq s) = k \ln(\underline{s}/s) + \epsilon \tag{4.2}$$

where k is the only unknown parameter. Therefore, we do not follow the usual practice in the literature of freely estimating an intercept. By doing that, we address the concerns of Clauset, Shalizi and Newman (2009) that regression lines are not valid distributions since, in our approach, $P(S \geq \underline{s}) = 1$, which is not generally valid if an intercept is freely estimated.[8] Besides this intercept restriction, this method is essentially a standard rank-size regression with binned data, as the number of firms with size $S \geq s$ equals the rank size of a firm with exactly s employees.[9]

Second, we use a maximum likelihood (ML) estimator. Virkar and Clauset (2014) show that an analytical solution for this ML estimator (MLE) can be obtained when the binning scheme is logarithmic. For arbitrary bins such as those of Table 4.1, however, a closed-form expression for this MLE does not exist, and thus, we obtain it numerically. Since it is computationally faster, we choose to solve

[7] Alternatively, we could apply OLS to a log-transformed histogram, gauging the power law exponent k from the empirical probability function instead of the empirical survival function. We choose not to follow this strategy because this estimator performed very poorly in Monte Carlo simulations (Clauset; Shalizi; Newman, 2009; Virkar; Clauset, 2014; Bottazzi; Pirino; Tamagni, 2015).

[8] Urzúa (2011, p.254) expresses similar concerns, arguing "the intercept is not a nuisance parameter in the regression."

[9] After all, if the j-th largest firm has size s, there must be j firms with size $S \geq s$ if we assign the highest possible rank to firms with the same size (e.g., if the two largest firms are the same size, we assign rank 2 for both).

the associated First-Order Condition (FOC), derived in Appendix 4.A.1, instead of directly maximizing the log-likelihood function as in Virkar and Clauset (2014).

If the correct $\underline{s}$ is chosen, it is known that OLS regression (4.2) consistently estimates k, since $\hat{P}(S \geq s)$ is a consistent estimator of $P(S \geq s)$ by the law of large numbers. It is also possible to show that MLEs for both binned and non-binned data are consistent and asymptotically efficient (Virkar; Clauset, 2014; Clauset; Shalizi; Newman, 2009).[10] But what about their small-sample performance? Clauset, Shalizi and Newman (2009), Virkar and Clauset (2014), and Bottazzi, Pirino and Tamagni (2015) study it through Monte Carlo exercises. They find OLS regression (4.2), but without the intercept constraint, is biased in small samples, although this bias is not typically very high.[11] MLEs have the best performance in binned data (Virkar; Clauset, 2014) and also in non-binned data (Clauset; Shalizi; Newman, 2009; Bottazzi; Pirino; Tamagni, 2015), accurately estimating k, with negligible bias.[12] These results are not unexpected as Aban and Meerschaert (2004) show that the MLE for non-binned data (with a small sample correction) is the best linear unbiased estimator (BLUE) and also the minimum variance unbiased estimator (MVUE).

Finally, regarding the lognormal distribution, we follow Virkar and Clauset (2014) and estimate its parameters μ and $\sigma > 0$ using only the MLE for binned data. As in the Pareto case, there is no analytic expression for this estimator, and thus, we obtain it by numerically solving the FOCs for the likelihood maximization. See Appendix 4.A.2 for the derivation of these FOCs.

4.2.2 Step 2: goodness-of-fit tests

Virkar and Clauset (2014) use a goodness-of-fit test to verify if a random variable follows an estimated distribution. This test requires a measure of the distance between empirical and estimated distributions. They suggest the Kolmogorov–Smirnov (KS) goodness-of-fit statistic, which can be formally defined

[10] The MLE for non-binned data is the known Hill (1975) estimator.

[11] With the (correct) intercept constraint, one should expect a more efficient estimation of k. See Schluter (2018) for proof of the rank-size regression case in large samples.

[12] For non-binned data, Bottazzi, Pirino and Tamagni (2015) also find very good performance for the OLS rank-size estimator with Gabaix and Ibragimov (2011) correction.

as

$$D = \max_{s \in \{\underline{s}, \ldots, 500\}} \left| \hat{P}(S < s) - P(S < s|\hat{\beta}) \right| = \max_{s \in \{\underline{s}, \ldots, 500\}} \left| \hat{P}(S \geq s) - P(S \geq s|\hat{\beta}) \right| \tag{4.3}$$

where $\hat{P}(\cdot)$ is the empirical probability and $P(\cdot|\hat{\beta})$ is the probability under an evaluated distribution with the estimated vector of parameters $\hat{\beta}$. Given the distance measure (4.3), an estimated distribution, and being n the number of firms with at least $\underline{s}$ employees, the p-value of the test can be computed following five steps:

1. Compute the distance D^* between estimated and empirical distributions using (4.3).
2. Generate a synthetic binned data set with n values that follows the same estimated distribution above $\underline{s}$.
3. Fit the model to this synthetic data set, obtaining a new estimated distribution.
4. From (4.3), compute the distance D between this new model and the synthetic data set.
5. Repeat steps 2–4 many times and report the fraction of the distances D that are at least as large as D^*.

Some comments are due. First, in the second step of this algorithm, Virkar and Clauset (2014) suggest the use of a semi-parametric bootstrap to generate a distribution that follows the estimated distribution above $\underline{s}$ and the empirical distribution below $\underline{s}$, which is necessary to them as they are also estimating $\underline{s}$. Since we are exogenously setting $\underline{s}$, we only need the distribution above $\underline{s}$. Second, they generate synthetic data above $\underline{s}$ by sampling from a non-binned distribution and then computing the synthetic bin counts. We choose to sample directly from a multinomial distribution whose events' probabilities are given by the probabilities of the bins, which can be easily computed from the estimated survival functions (see Appendix 4.A). Third, we generate 10,000 synthetic data sets for each test, which is probably high enough as Virkar and Clauset (2014) show that with 2,500 simulations, one can gauge the p-value to within 0.01 of the true value. Fourth, we compute the test for Pareto and lognormal distributions, for each considered estimator. We also test a strong Zipf's law, when no estimation is required as it is a Pareto distribution with $k = 1$.

4.2.3 Step 3: comparing the distributions

Virkar and Clauset (2014) suggest the use of the likelihood ratio test proposed by Vuong (1989) to compare non-nested distributions in binned data. Suppose one wants to compare distribution A against distribution B, which are not nested. Let $\mathcal{L}_d = \prod_{i=j}^{m} (p_{d,i})^{h_i}$ be the likelihood of distribution $d = A, B$, where $p_{d,i}$ is the probability that some observation falls within the i-th bin under distribution d and h_i is the number of raw observations in the i-th bin. Note that there are m bins, but the distributions hold only from the j-th bin, meaning $\underline{s}$ is the lower bound of the j-th bin. Given that, the log-likelihood ratio of comparing A against B is $\mathcal{R} \equiv \ln \mathcal{L}_A - \ln \mathcal{L}_B$. Let us also define the normalized log-likelihood ratio as $\mathcal{R}_n \equiv \mathcal{R}/\sqrt{2n\hat{\sigma}^2_{\mathcal{R}}}$, where $\hat{\sigma}^2_{\mathcal{R}}$ is the estimated variance on the log-likelihood ratio $\mathcal{R}$, that is,

$$\hat{\sigma}^2_{\mathcal{R}} \equiv \frac{1}{n} \sum_{i=j}^{m} h_i \left[(\ln p_{A,i} - \ln p_{B,i}) - \mathcal{R}/n \right]^2 \tag{4.4}$$

$n \equiv \sum_{i=j}^{m} h_i$ is the number of firms with at least $\underline{s}$ employees or, equivalently, the number of firms at the j-th bin or above. Vuong (1989) shows that under the null that the two distributions are equivalent, $\sqrt{2}\mathcal{R}_n \xrightarrow{D} N(0,1)$; under the alternative that distribution A is better, $\sqrt{2}\mathcal{R}_n \xrightarrow{a.s.} +\infty$; finally, under the alternative that distribution B is better, $\sqrt{2}\mathcal{R}_n \xrightarrow{a.s.} -\infty$. As a consequence, under the null hypothesis, in large samples,

$$\begin{aligned} P\left(|\mathcal{R}| \geq |\mathcal{R}^*|\right) =& P\left(\sqrt{2}|\mathcal{R}_n| \geq \sqrt{2}|\mathcal{R}^*_n|\right) = 2 \times P\left(\sqrt{2}\mathcal{R}_n \geq \sqrt{2}|\mathcal{R}^*_n|\right) \\ P\left(|\mathcal{R}| \geq |\mathcal{R}^*|\right) =& 2\left\{1 - (1/2)\left[1 + \text{erf}\left(\sqrt{2}|\mathcal{R}^*_n|/\sqrt{2}\right)\right]\right\} = 1 - \text{erf}\left(|\mathcal{R}^*_n|\right) \end{aligned} \tag{4.5}$$

where $\text{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ is the Gaussian error function. Hence, setting a significance level p^*, one can get $T > 0$ that solves $p^* = 1 - \text{erf}(T)$. If $\mathcal{R}_n \geq T$ ($\mathcal{R}_n \leq -T$), the null is rejected in favor of A being better (worse) than B, while the null is not rejected if $-T < \mathcal{R}_n < T$.

We apply this test to compare (i) Pareto against lognormal and (ii) strong Zipf's law against lognormal, using Pareto and lognormal densities as estimated by ML. Testing strong Zipf's law against the Pareto distribution is equivalent to verifying if $k = 1$. However, standard OLS t-tests would not be reliable here since they have a strong tendency to over-reject the null $k = 1$, as Gabaix and Ibragimov (2011) and Bottazzi, Pirino and Tamagni (2015) show through Monte Carlo exercises. Indeed, when sampling from a Zipf distribution, Bottazzi, Pirino and Tamagni (2015) could reject the null $k = 1$ at 5% confidence level 60 – 70% of the time! Given that, we follow Virkar and Clauset (2014) and verify it using the ML estimates and a standard likelihood ratio test. Under the null $k = 1$, it is known that $2|\mathcal{R}|$ is asymptotically chi-squared distributed with one degree of freedom, where $\mathcal{R}$ is the log-likelihood ratio between Pareto and Zipf distributions.

4.3 Results

4.3.1 Step 1: fitting the distributions

Figures 4.1 and 4.2 plot empirical and estimated survival functions in 1996 and 2020, respectively, our sample's initial and final years. In both cases, we present the results at economy-wide and industry levels, with $\underline{s} = 5, 20, 50$. The axes of each plot are in logarithmic scale, with $P(S \geq s)$ in the vertical axis and s in the horizontal axis, implying estimated survival functions are straight lines in Pareto cases. Inside each plot, we show the OLS/ML estimates of k, $\hat{k}$, and the (centered) R^2 for each estimator/distribution computed from these plotted data. As can be seen, both distributions fit the data well. For $\underline{s} = 5$, the Pareto distribution does a better job, especially closer to the upper tail, while for $\underline{s} = 20$ and mainly for $\underline{s} = 50$, both distributions fit similarly well. Focusing on the Pareto case, note estimates of k are relatively robust to the choice of estimator, particularly for higher $\underline{s}$. Additionally, all estimates of k are around one and typically become closer to this level as $\underline{s}$ increases.

These results are not specific to 1996 and 2020 or $\underline{s} = 5, 20, 50$. In Figures 4.3 and 4.4, we plot the (centered) R^2 for each year, industry, lower bound $\underline{s}$, and estimator/distribution for 1996-2006 and 2006-2020, respectively. The fit of each model is very good for $\underline{s} = 20, 30, 50$, while for $\underline{s} = 5$, the lognormal fit is usually worse. Moreover, especially for $\underline{s} = 10, 20$, the ML Pareto estimate has the worst fit for the services sector. The power law exponent k estimates for 1996-2006 and 2006-2020 are shown in Figures 4.5 and 4.6, respectively. Several things are worth noting about these estimates. First, they are around one, typically between 0.8 and 1.2, and approach the unitary value for higher values of $\underline{s}$. Second, they are also surprisingly stable over time. Fujiwara et al. (2004) find similar stability for the UK, France, Italy, and Spain between 1993 and 2001, with firm size measured by total assets, number of employees, and sales (except for the UK). Resende and Cardoso (2022), using net revenue to measure firm size, also estimate a relatively stable power law exponent for Brazil between 1999 and 2019. Third, estimates by OLS vary much less than those by ML when a different $\underline{s}$ is chosen, similar to what Aban and Meerschaert (2004) find for the daily trading volume of Amazon, Inc. stock.

4.3.2 Step 2: goodness-of-fit tests

The computed p-values for the goodness-of-fit tests are shown in Figures 4.7 and 4.8 for 1996-2006 and 2006-2020, respectively, which are essentially the same for ML and OLS estimators of the Pareto density. Besides the very good fit of the estimated distributions shown previously, these tests reject both Pareto and lognormal distributions in most cases. Consistent with these findings, Resende (2004) does not find strong evidence supporting a lognormal distribution of firm size by the number of employees in Brazil either. Similarly, the Zipf distribution is also usually rejected. For Pareto and Zipf distributions, the main exception to these conclusions is agriculture, particularly for $\underline{s} = 20, 30, 50$ when the Pareto distribution is not rejected in almost all years, and the strong Zipf's law cannot be rejected for several years between 2006 and 2020. In the lognormal case, the main exception is $\underline{s} = 50$, when the distribution is typically not rejected (except for industry between 2006 and 2020).

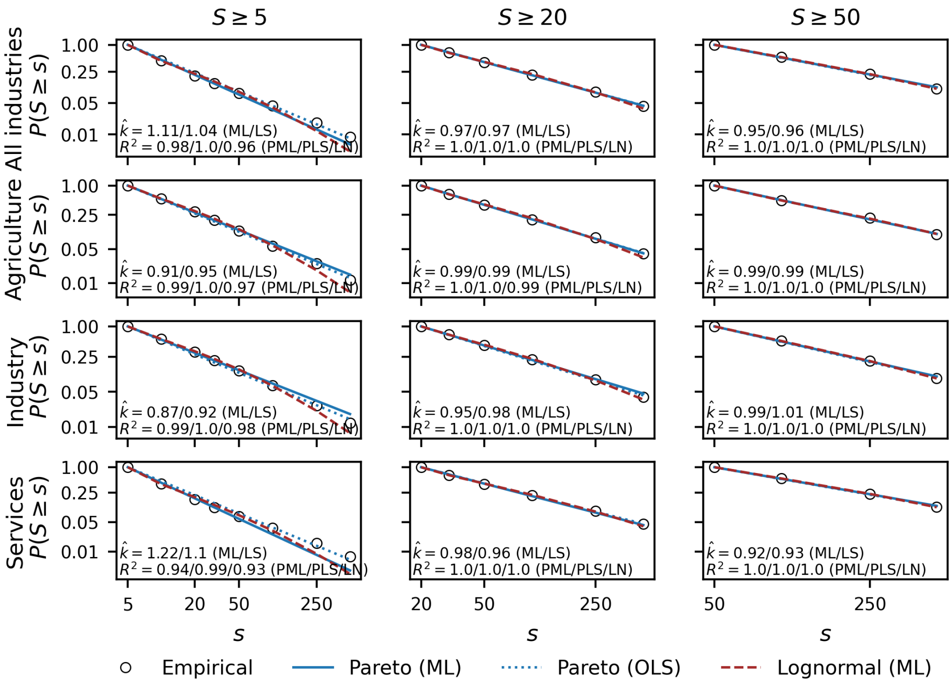


Figure 4.1 – Models fit in 1996 (axes in logarithmic scale).

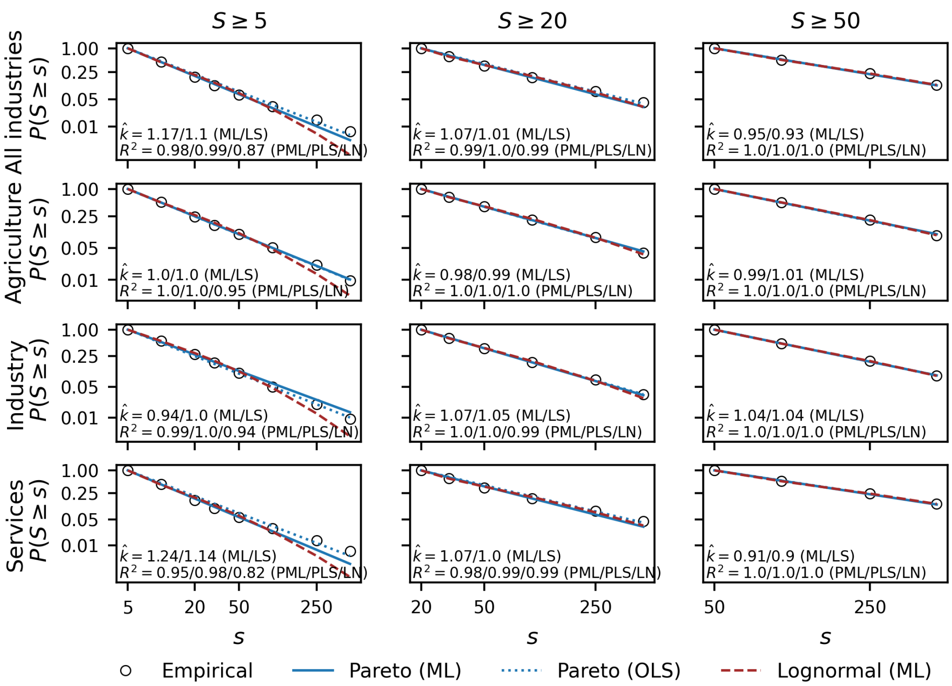


Figure 4.2 – Models fit in 2020 (axes in logarithmic scale).

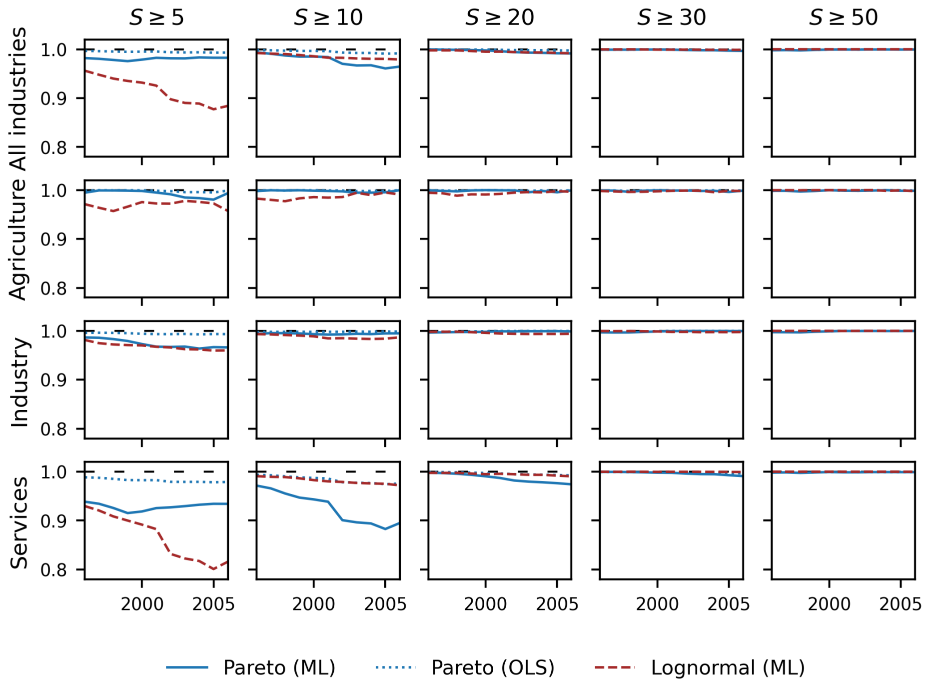


Figure 4.3 – Centered R^2, 1996-2006.

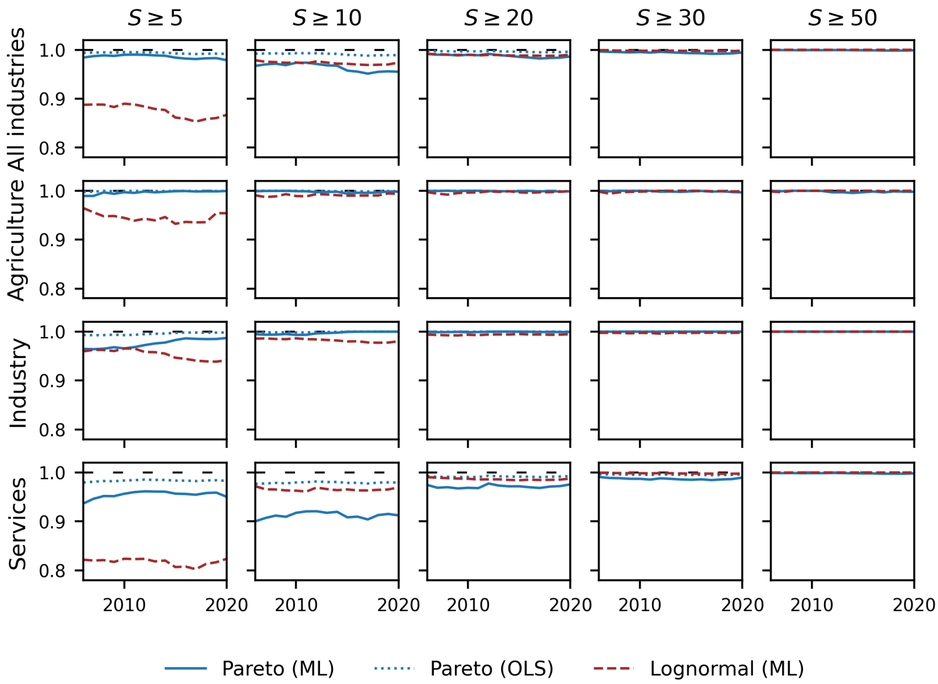


Figure 4.4 – Centered R^2, 2006-2020.

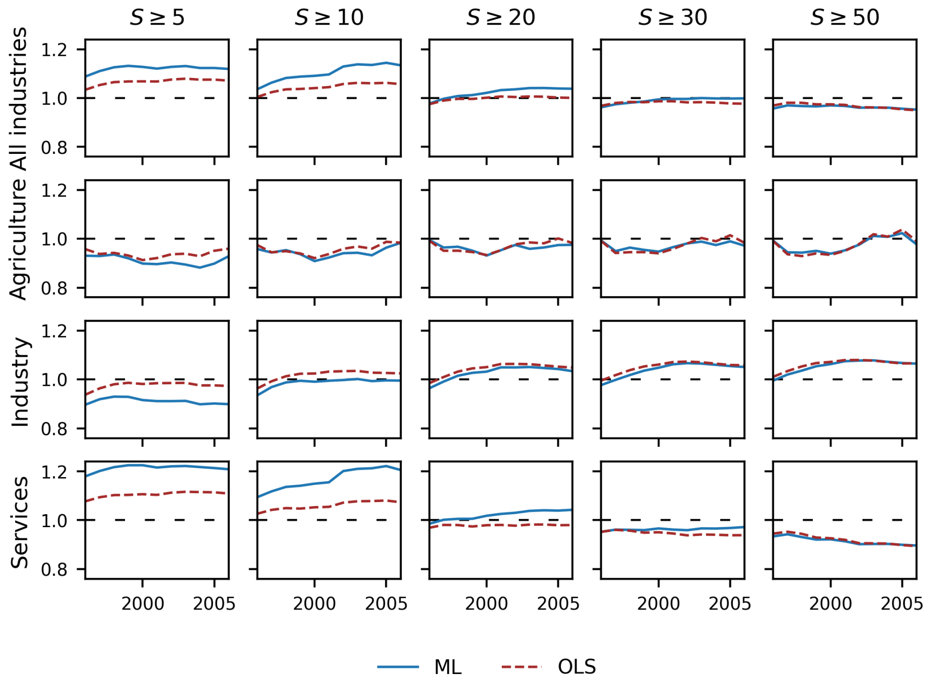


Figure 4.5 – k estimates, 1996-2006.

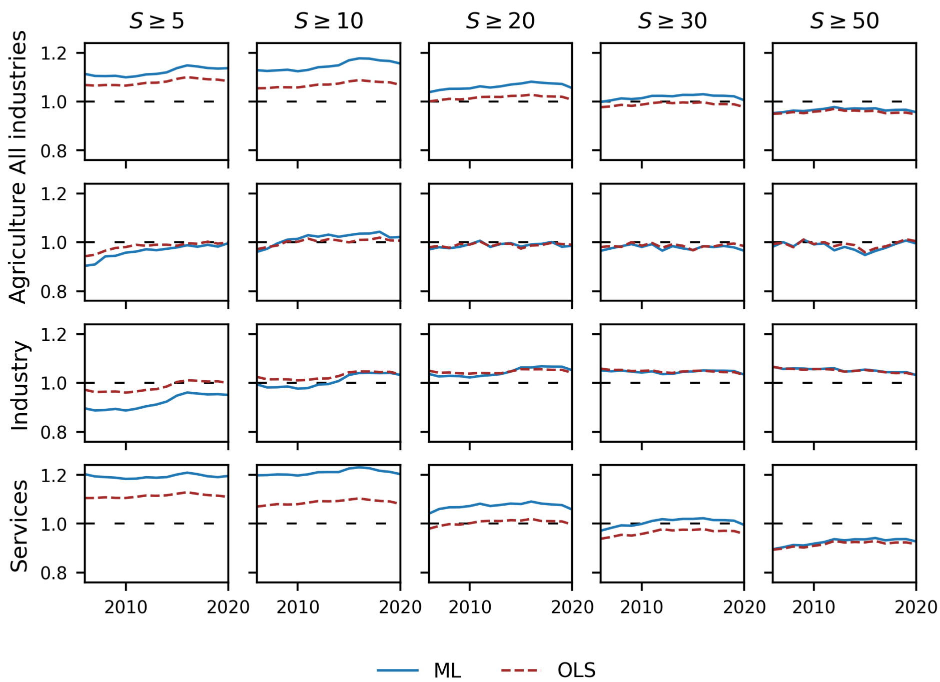


Figure 4.6 – k estimates, 2006-2020.

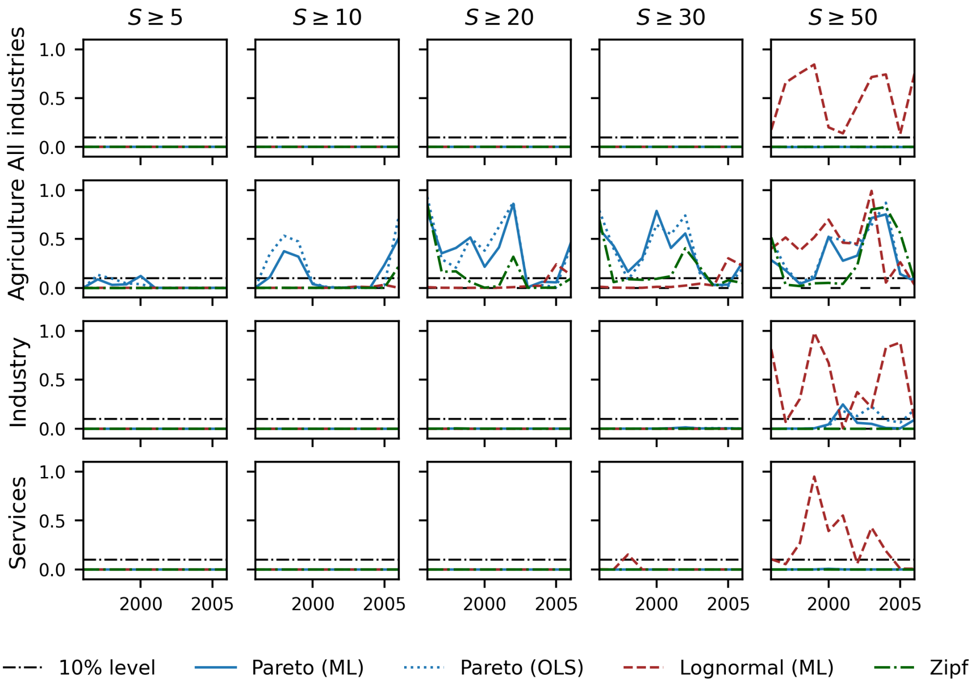


Figure 4.7 – p-value of the goodness-of-fit test, 1996-2006.

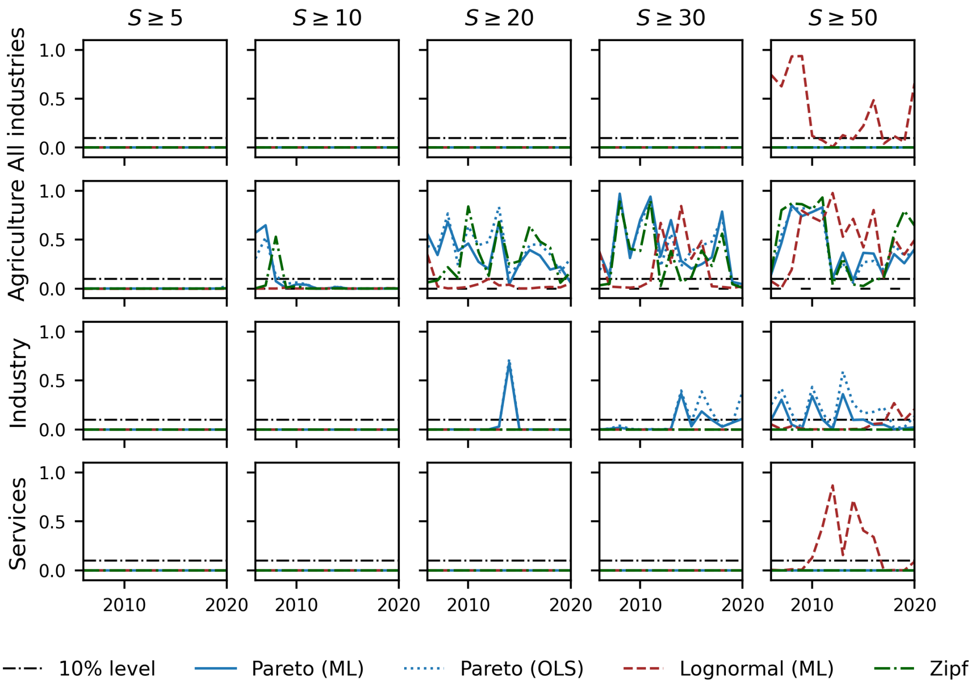


Figure 4.8 – p-value of the goodness-of-fit test, 2006-2020.

4.3.3 Step 3: comparing the distributions

In Figures 4.9 and 4.10, we plot the normalized log-likelihood ratio $\mathcal{R}_n$ and the thresholds at 10% level for 1996-2006 and 2006-2020, respectively. To make it easier to visualize the results, we plot $\mathcal{R}_n = 2$ ($\mathcal{R}_n = -2$) when $\mathcal{R}_n \geq 2$ ($\mathcal{R}_n \leq -2$). The results confirm that the lognormal provides a strong test for the Pareto distribution since there is no single winner between them in all cases. Typically, the Pareto distribution beats the lognormal for lower $\underline{s}$, while the lognormal wins for higher $\underline{s}$, particularly for $\underline{s} = 50$, which is consistent with the goodness-of-fit tests results seen in the last section. Furthermore, it is worth mentioning that several of these results are consistent with the R^2 shown in Figures 4.3 and 4.4. For instance, both likelihood and R^2 of the Pareto distribution are mostly higher for $\underline{s} = 5$ but lower in the services sector for $\underline{s} = 10, 20, 30, 50$. Finally, when comparing strong Zipf's law and lognormal, the latter rarely loses. The main exception is industry under $\underline{s} = 10$.

The p-values of testing strong Zipf's law against the Pareto distribution for 1996-2006 and 2006-2020 are shown in Figures 4.11 and 4.12, respectively. In almost all industries and years, we can reject $k = 1$. The main exception is 2006-2020 agriculture under $\underline{s} = 20, 30, 50$. Therefore, although the estimates of the power law exponent k are around one, especially for higher $\underline{s}$, they are not *exactly* one in most cases.

4.3.4 Discussion

Let us summarize and discuss our findings from all three steps. Although a Zipf distribution can be ruled out, we estimate power exponent $k \approx 1$ with good data fit, especially for higher $\underline{s}$, consistent with Zipf's law. However, a lognormal density also performs well and even outperforms the Pareto distribution in certain cases. The main issue is that the goodness-of-fit tests ruled out that the firm size distribution in Brazil is *exactly* Pareto, Zipf, or lognormal in most cases. Nevertheless, as Gabaix (2009, p.285) points out,

> With an infinitely large empirical data set, one can reject any nontautological theory. Hence, the main question of empirical work should be

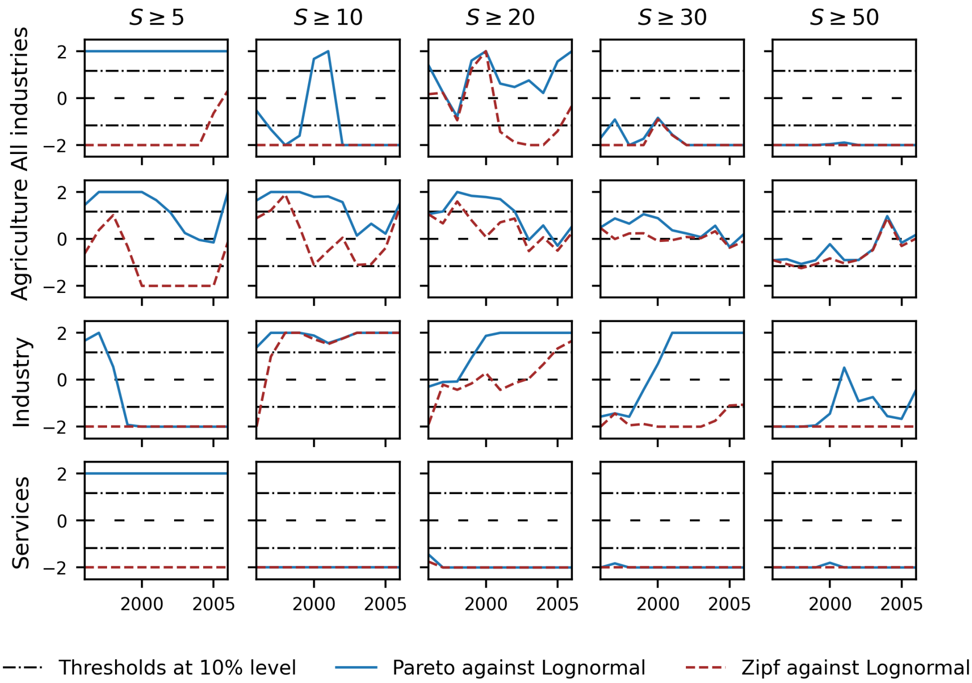


Figure 4.9 – Normalized log-likelihood ratio, 1996-2006.

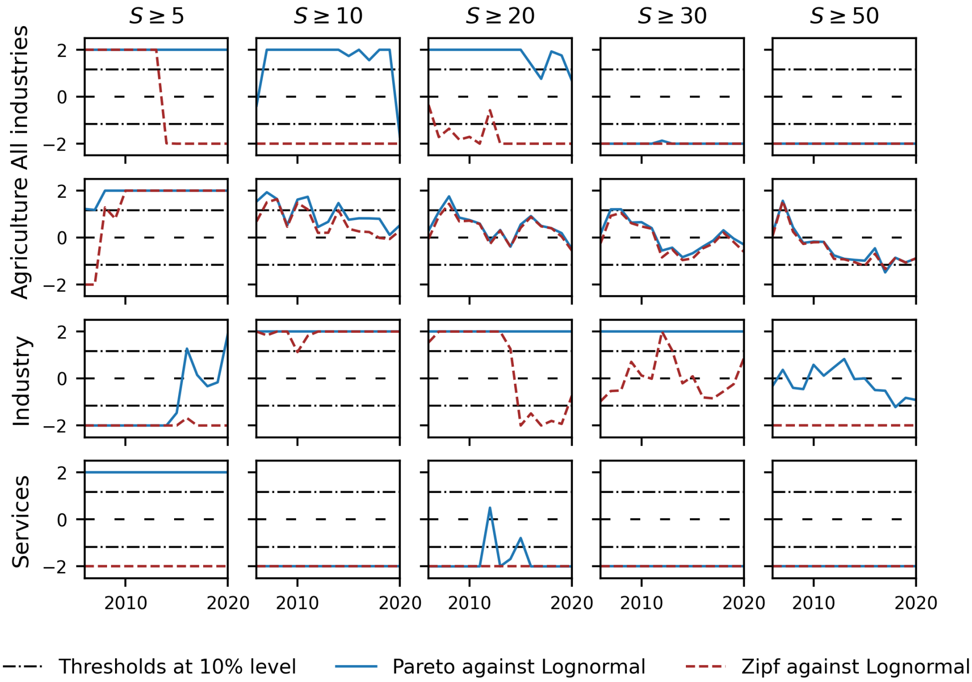


Figure 4.10 – Normalized log-likelihood ratio, 2006-2020.

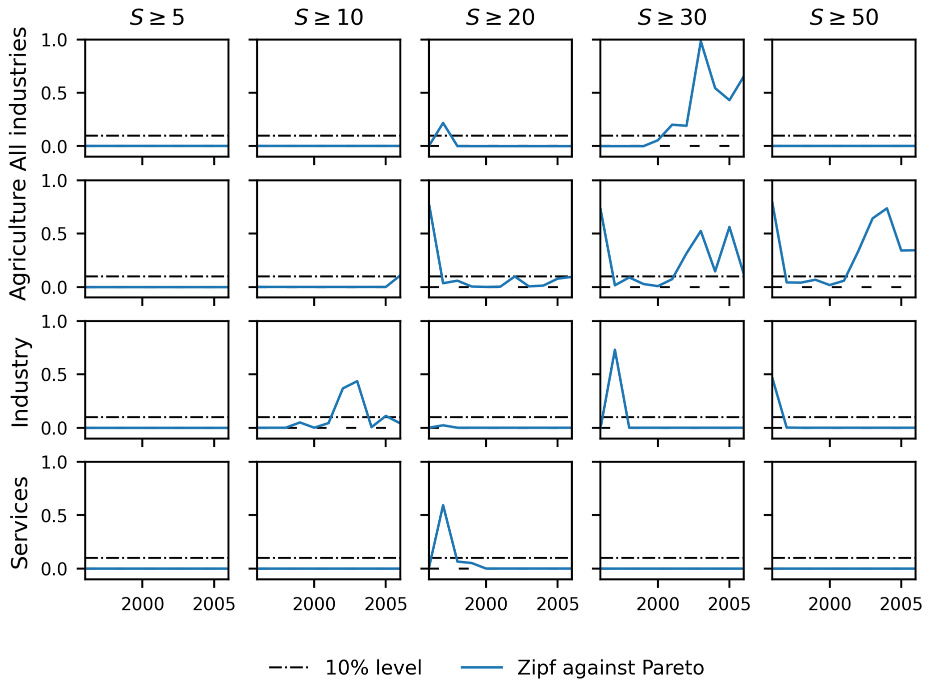


Figure 4.11 – p-value of the standard likelihood ratio test, 1996-2006.

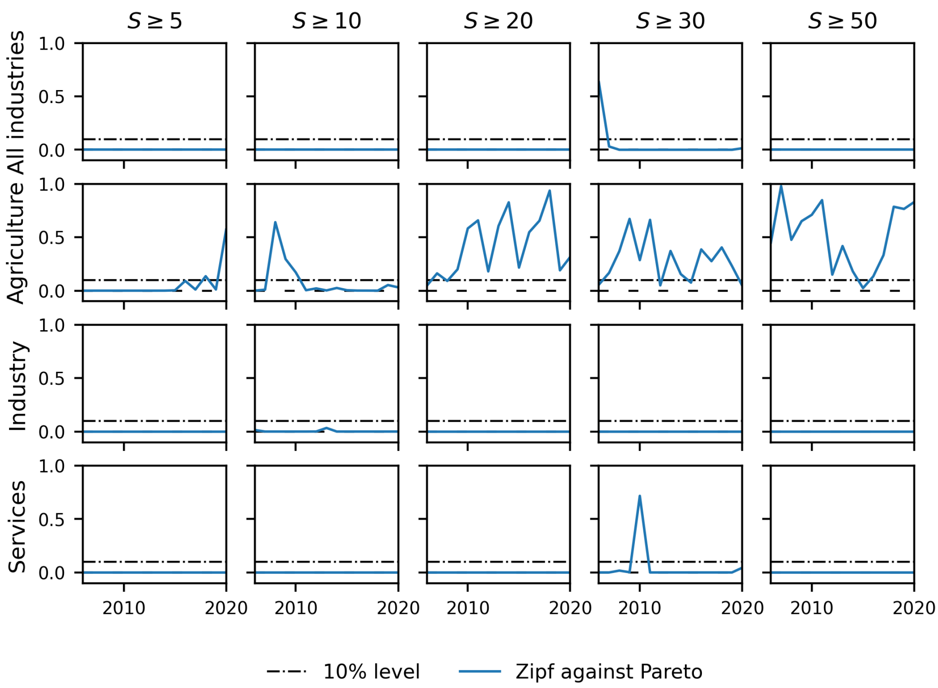


Figure 4.12 – p-value of the standard likelihood ratio test, 2006-2020.

> how well a theory fits, rather than whether it fits perfectly (i.e., within the standard errors). [...] Consistent with these suggestions, some of the debate on Zipf's law should be cast in terms of how well, or poorly, it fits, rather than whether it can be rejected.

From that point of view, Pareto and lognormal distributions are still useful benchmarks as they provide very good, although not perfect, approximation to data. This can be seen more clearly in Table 4.2, which shows empirical and estimated bins' probabilities over the support $S \geq 20$ for ML Pareto and lognormal distributions in 1996 and 2020. These good fits hold at the economy-wide level and also for agriculture, industry, and services alone, for each year between 1996 and 2020. As it is well known, Brazil experienced an economic boom in the 2000s and a bust with huge volatility in the 2010s, which possibly explains why the total number of firms varied so much over time (Figure 4.13), but firm size distribution remained basically unchanged throughout the entire period, always close to Zipf's law. This is a rather remarkable result even if this "law" is not *exactly* valid, since, as Gabaix (2009, p.256) points out for the distribution of city size, "there is no tautology causing the data to automatically generate this shape."

4.4 Conclusion

In this paper, we evaluate Zipf's law for the distribution of firm size by the number of employees in Brazil. Remarkably, we find that Zipf's law provides a very good, although not perfect, approximation to data for each year between 1996 and 2020 at the economy-wide level and also for agriculture, industry, and services alone. However, a lognormal distribution also performs well and even outperforms Zipf's law in certain cases.

Our analyses are based on publicly available data from CEMPRE, which facilitates other researchers' reproduction and exploration of our results. Nevertheless, this choice also has relevant shortcomings due to binning, suggesting working with CEMPRE firm-level data may be an interesting avenue for future research. First, binning leads to a loss of information, such that a higher number of sampled firms is required to achieve the same accuracy in estimating and testing the distributions

Table 4.2 – Empirical and ML estimated bins' probabilities over $S \geq 20$

Number of employees	1996 Empirical	1996 Pareto	1996 Lognormal	2020 Empirical	2020 Pareto	2020 Lognormal
All industries						
20 to 29	33.6	32.6	34.1	38.0	35.2	38.9
30 to 49	25.8	26.4	24.2	26.5	27.3	23.8
50 to 99	19.1	20.1	19.4	17.6	19.7	17.9
100 to 249	12.6	12.3	13.6	9.9	11.2	12.0
250 to 499	4.7	4.2	4.9	4.0	3.5	4.2
500 or more	4.3	4.4	3.8	4.1	3.2	3.2
Agriculture						
20 to 29	33.4	33.2	34.0	34.2	32.9	34.3
30 to 49	26.5	26.6	24.9	24.4	26.5	24.4
50 to 99	20.3	20.0	19.7	20.5	20.1	19.4
100 to 249	11.2	12.1	13.4	12.2	12.2	13.4
250 to 499	4.7	4.0	4.7	4.8	4.1	4.8
500 or more	4.0	4.1	3.4	3.9	4.2	3.7
Industry						
20 to 29	31.2	32.1	31.6	36.1	35.1	36.6
30 to 49	26.2	26.2	25.2	26.4	27.3	25.1
50 to 99	20.7	20.2	20.6	19.3	19.7	19.1
100 to 249	13.2	12.5	14.2	10.9	11.2	12.3
250 to 499	4.8	4.3	4.9	3.8	3.5	4.1
500 or more	3.9	4.6	3.5	3.4	3.2	2.8
Services						
20 to 29	34.9	32.8	35.5	38.8	35.3	39.8
30 to 49	25.5	26.5	23.6	26.5	27.3	23.2
50 to 99	18.1	20.1	18.8	16.9	19.6	17.5
100 to 249	12.3	12.2	13.2	9.5	11.1	11.8
250 to 499	4.7	4.1	4.9	4.0	3.5	4.3
500 or more	4.5	4.2	4.0	4.4	3.2	3.4

when data is binned (Virkar; Clauset, 2014). One may argue that this information loss could be especially harsh in the CEMPRE database since there is little information on the upper tail. After all, the last bin available contains firms with 500 or more employees, which is probably too wide since the biggest firms would typically have a much larger number of employees. In any case, since our samples are large, this may not be such a severe problem here. Second, one can easily explore more flexible distributions when working with non-binned data. Kondo,

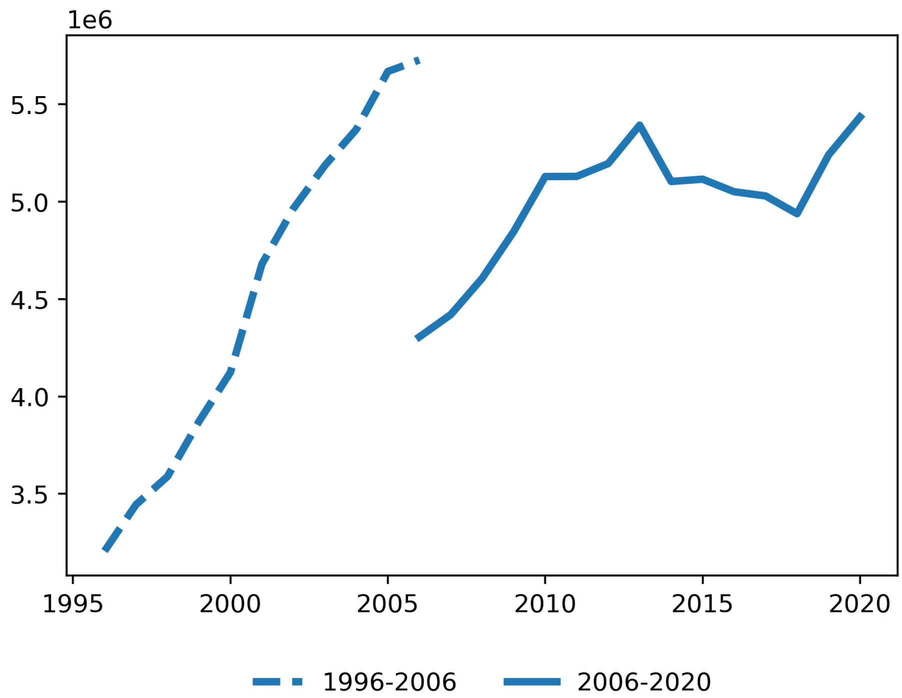


Figure 4.13 – Total number of firms.

Lewis and Stella (2023) estimate statistical mixtures and convolutions of Pareto and lognormal distributions in the US, finding these combinations significantly beat each distribution alone. Alternatively, one can apply Lagrange multiplier tests, verifying the null of power or Zipf's law against a distribution that *nests* the Pareto density. One advantage of these tests is that they do not require the estimation of the more general density, which may be challenging in some cases. In principle, such tests could be applied to binned data; however, to the best of our knowledge, so far, they have been developed only for non-binned data (Urzúa, 2000; Goerlich, 2013; Urzúa, 2020). Resende and Cardoso (2022) apply these tests to the distribution of firm size by net revenue in Brazil. They consider the 1,000, 500, and 100 largest firms between 1999 and 2019, finding strong support for power or Zipf's law only (i) against the Pareto type II distribution and (ii) among the 100 largest firms. This suggests investigating distributions that nest the Pareto density using CEMPRE firm-level data can be fruitful.

Appendix of Chapter 4

Following a notation similar to Virkar and Clauset (2014), let $B = \{b_1, ..., b_m\}$ be a set of bin boundaries, $b_1 = 0$, $b_i > 0$ for $i \in \{2, ..., m\}$, and $b_j > b_i$ for $j > i$ and $i, j \in \{1, 2, ..., m\}$. With these boundaries, we define m bins, with $[b_i, b_{i+1})$ being the i-th bin, $i \in \{1, 2, ..., m-1\}$, and $[b_m, +\infty)$ being the m-th bin. Denote by $H \in \{h_1, h_2, ..., h_m\}$ the set of bin counts, such that h_i is the number of raw observations in the i-th bin, $i = 1, 2, ..., m$. Lastly, let $n \equiv \sum_{i=j}^{m} h_i$ be the number of firms with at least $b_j = \underline{s} > 0$ employees.

4.A Maximum likelihood estimator for binned data

Suppose S, $S \geq \underline{s} = b_j > 0$, follows a certain distribution. Given that, the log-likelihood function for the binned data over this support is

$$\begin{aligned}\mathcal{L} =& \ln\left[P(S \geq b_m)^{h_m} P(\underline{s} \leq S < b_{j+1})^{h_j} \prod_{i=j+1}^{m-1} P(b_i \leq S < b_{i+1})^{h_i}\right] \\ \mathcal{L} =& h_m \ln P(S \geq b_m) + h_j \ln\left[1 - P(S \geq b_{j+1})\right] + \sum_{i=j+1}^{m-1} h_i \ln P(b_i \leq S < b_{i+1})\end{aligned} \tag{4.A.1}$$

which allows us to get the maximum likelihood estimator (MLE) of the distributional parameters. One possibility is to numerically maximize the log-likelihood function (4.A.1). Nevertheless, a computationally faster way is to derive and numerically solve the associated First-Order Conditions (FOCs), derived for Pareto and lognormal distributions in the following.

4.A.1 Pareto distribution

If S, $S \geq \underline{s} = b_j > 0$, is Pareto distributed with shape parameter $k > 0$,

$$\begin{aligned}P(S \geq b_i) =& \left(\underline{s}/b_i\right)^k, \; i = j, j+1, ..., m \\ P(b_i \leq S < b_{i+1}) =& P(S \geq b_i) - P(S \geq b_{i+1}) = \underline{s}^k \left(b_i^{-k} - b_{i+1}^{-k}\right), \; i = j, ..., m-1\end{aligned}$$

Plugging these probabilities into the log-likelihood function (4.A.1),

$$\mathcal{L} = h_m k \ln \underline{s} - h_m k \ln b_m + \sum_{i=j}^{m-1} \left[h_i k \ln \underline{s} + h_i \ln \left(b_i^{-k} - b_{i+1}^{-k} \right) \right]$$

$$\mathcal{L} = nk \ln \underline{s} - h_m k \ln b_m + \sum_{i=j}^{m-1} h_i \ln \left(b_i^{-k} - b_{i+1}^{-k} \right) \tag{4.A.2}$$

From (4.A.2), which is equivalent to equation (3.1) of Virkar and Clauset (2014), one can obtain the desired FOC:

$$\frac{\partial \mathcal{L}}{\partial k} = n \ln \underline{s} - h_m \ln b_m - \sum_{i=j}^{m-1} h_i \left[\frac{b_i^{-k} \ln b_i - b_{i+1}^{-k} \ln b_{i+1}}{b_i^{-k} - b_{i+1}^{-k}} \right] = 0 \tag{4.A.3}$$

4.A.2 Lognormal distribution

If $S - \underline{s} = S - b_j$, $S - b_j > 0$, is lognormally distributed with parameters μ and $\sigma > 0$,

$$P(S \geq b_i) = \frac{1 - \text{erf}\,(z_i)}{2},\ i = j+1, j+2, ..., m$$

$$P(b_i \leq S < b_{i+1}) = \frac{\text{erf}\,(z_{i+1}) - \text{erf}\,(z_i)}{2},\ i = j+1, ..., m-1$$

where $z_i \equiv \frac{\ln(b_i - \underline{s}) - \mu}{\sigma\sqrt{2}}$ and $\text{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ is the Gaussian error function. Plugging these probabilities into the log-likelihood function (4.A.1),

$$\mathcal{L} = h_m \ln \left[\frac{1 - \text{erf}\,(z_m)}{2} \right] + h_j \ln \left[\frac{1 + \text{erf}\,(z_{j+1})}{2} \right] + \sum_{i=j+1}^{m-1} h_i \ln \left[\frac{\text{erf}\,(z_{i+1}) - \text{erf}\,(z_i)}{2} \right] \tag{4.A.4}$$

The FOCs for the maximization of the log-likelihood function (4.A.4) are

$$\frac{\partial \mathcal{L}}{\partial \mu} = h_m \frac{\text{erf}'(z_m)}{P(S \geq b_m) 2\sigma\sqrt{2}} - h_j \frac{\text{erf}'(z_{j+1})}{P(S < b_{j+1}) 2\sigma\sqrt{2}} + \sum_{i=j+1}^{m-1} h_i \frac{\text{erf}'(z_i) - \text{erf}'(z_{i+1})}{P(b_i \leq S < b_{i+1}) 2\sigma\sqrt{2}}$$

$$\therefore 0 = h_m \frac{e^{-z_m^2}}{P(S \geq b_m)} - h_j \frac{e^{-z_{j+1}^2}}{P(S < b_{j+1})} + \sum_{i=j+1}^{m-1} h_i \frac{e^{-z_i^2} - e^{-z_{i+1}^2}}{P(b_i \leq S < b_{i+1})} \tag{4.A.5}$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = h_m \frac{\text{erf}'(z_m) z_m}{P(S \geq b_m) 2\sigma} - h_j \frac{\text{erf}'(z_{j+1}) z_{j+1}}{P(S < b_{j+1}) 2\sigma} + \sum_{i=j+1}^{m-1} h_i \frac{\text{erf}'(z_i) z_i - \text{erf}'(z_{i+1}) z_{i+1}}{P(b_i \leq S < b_{i+1}) 2\sigma}$$

$$\therefore 0 = h_m \frac{z_m e^{-z_m^2}}{P(S \geq b_m)} - h_j \frac{z_{j+1} e^{-z_{j+1}^2}}{P(S < b_{j+1})} + \sum_{i=j+1}^{m-1} h_i \frac{z_i e^{-z_i^2} - z_{i+1} e^{-z_{i+1}^2}}{P(b_i \leq S < b_{i+1})} \tag{4.A.6}$$

where we use $\text{erf}'(z) = \frac{2}{\sqrt{\pi}} e^{-z^2}$ to get each condition.

5 Conclusions

In the first paper, using a Cournot model, we decompose TFP into technology and allocative efficiency components from 1950 to 2019 for up to a hundred countries from the Penn World Table 10.01. This decomposition enables a reexamination of key facts of economic growth. Our evaluation of the world income frontier, proxied by the US, reveals that changes in misallocation can significantly impact short-run growth. For example, during 2000-2007, the US witnessed notable technological improvement coupled with declining allocative efficiency, suggesting that the dot-com boom and advancements in IT led to productivity gains but concentrated in certain firms. On a more general note, the technology component seems to grow more steadily than the TFP itself, around 1% per year. Notable exceptions are the periods of 1954-1973 and 2000-2007 when technology contributed approximately 2% annually. Turning to a global perspective, our analysis suggests that misallocation plays a significant role in explaining cross-country income differences, even though a considerable unexplained portion persists. We also find a lack of convergence in allocative efficiency, suggesting market-power-driven misallocation is linked, in the long run, to long-lasting country-specific factors such as institutions.

The second paper uses the Cournot model developed in the first article to decompose the Brazilian TFP between 2000 and 2019. We find an overall improvement in allocative efficiency, reflecting the observed increase in the labor income share and, thus, the estimated decrease in the average markup. We also find that TFP cycles are essentially due to allocative efficiency, with the economic boom in the mid-2000s being primarily attributed to efficiency gains. The technology frontier grows much more steadily, suggesting this reflects the structural characteristics of the economy. Therefore, since allocative improvements could not occur indefinitely, the annual technology growth found, around 0.8-0.9%, can be seen as the current structural, long-run, growth level of Brazilian TFP.

Finally, the third paper evaluates Zipf's law for the distribution of firm size by the number of employees in Brazil, using publicly available binned annual data from CEMPRE, which covers all formal organizations. Remarkably, we find that this “law” provides a very good, although not perfect, approximation to data for each year between 1996 and 2020 at the economy-wide level and also for agriculture, industry, and services alone. However, a lognormal distribution also performs well and even outperforms Zipf’s law in certain cases.

Bibliography

Aban, I. B.; Meerschaert, M. M. Generalized least-squares estimators for the thickness of heavy tails. *Journal of Statistical Planning and Inference*, Elsevier, v. 119, n. 2, p. 341–352, 2004.

Ackerberg, D. A.; Caves, K.; Frazer, G. Identification properties of recent production function estimators. *Econometrica*, Wiley Online Library, v. 83, n. 6, p. 2411–2451, 2015.

Atkeson, A.; Burstein, A. Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, v. 98, n. 5, p. 1998–2031, 2008.

Autor, D.; Dorn, D.; Katz, L. F.; Patterson, C.; Reenen, J. V. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, Oxford University Press, v. 135, n. 2, p. 645–709, 2020.

Axtell, R. L. Zipf distribution of U.S. firm sizes. *Science*, American Association for the Advancement of Science, v. 293, n. 5536, p. 1818–1820, 2001.

Baqaee, D. R.; Farhi, E. Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, Oxford University Press, v. 135, n. 1, p. 105–163, 2020.

Barkai, S. Declining labor and capital shares. *The Journal of Finance*, Wiley Online Library, v. 75, n. 5, p. 2421–2463, 2020.

Basu, S.; Fernald, J.; Fisher, J.; Kimball, M. Sector-specific technical change. *Manuscript, Federal Reserve Bank of San Francisco*, 2013.

Basu, S.; Fernald, J. G. Aggregate productivity and aggregate technology. *European Economic Review*, Elsevier, v. 46, n. 6, p. 963–991, 2002.

Basu, S.; Fernald, J. G.; Kimball, M. S. Are technology improvements contractionary? *American Economic Review*, American Economic Association, v. 96, n. 5, p. 1418–1448, 2006.

Benkard, C. L.; Yurukoglu, A.; Zhang, A. L. Concentration in product markets. *NBER Working Paper Series*, n. 28745, 2021.

Bergeaud, A.; Cette, G.; Lecat, R. The role of production factor quality and technology diffusion in twentieth-century productivity growth. *Cliometrica*, Springer, v. 12, n. 1, p. 61–97, 2018.

Berger, D.; Herkenhoff, K.; Mongey, S. Labor market power. *American Economic Review*, v. 112, n. 4, p. 1147–93, 2022.

Bernard, A. B.; Eaton, J.; Jensen, J. B.; Kortum, S. Plants and productivity in international trade. *American Economic Review*, v. 93, n. 4, p. 1268–1290, 2003.

Berry, S.; Gaynor, M.; Morton, F. S. Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives*, v. 33, n. 3, p. 44–68, 2019.

Blundell, R.; Bond, S. Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews*, Taylor & Francis, v. 19, n. 3, p. 321–340, 2000.

Bottazzi, G.; Pirino, D.; Tamagni, F. Zipf law and the firm size distribution: a critical discussion of popular estimators. *Journal of Evolutionary Economics*, Springer, v. 25, n. 3, p. 585–610, 2015.

Buddana, A.; Kozubowski, T. J. Discrete Pareto distributions. *Economic Quality Control*, De Gruyter, v. 29, n. 2, p. 143–156, 2014.

Busso, M.; Madrigal, L.; Pagés, C. Productivity and resource misallocation in Latin America. *The BE Journal of Macroeconomics*, De Gruyter, v. 13, n. 1, p. 903–932, 2013.

Byrne, D. M.; Oliner, S. D.; Sichel, D. E. Is the information technology revolution over? *International Productivity Monitor*, v. 25, p. 20–36, Spring 2013.

Calligaris, S.; Criscuolo, C.; Marcolin, L. Mark-ups in the digital era. *OECD Science, Technology and Industry Working Papers*, OECD Publishing, n. 2018/10, 2018.

Carvalho, A.; Santos, T. T. O. Is the equity risk premium compressed in Brazil? *Central Bank of Brazil Working Paper Series*, n. 527, 2020.

Caselli, F. Accounting for cross-country income differences. In: Aghion, P.; Durlauf, S. (Ed.). *Handbook of Economic Growth.* : Elsevier, 2005. v. 1A, Chap. 9, p. 679–741.

Caselli, F. Accounting for cross-country income differences: Ten years later. *World Bank Background Paper*, World Bank, 2016.

Clauset, A.; Shalizi, C. R.; Newman, M. E. Power-law distributions in empirical data. *SIAM Review*, SIAM, v. 51, n. 4, p. 661–703, 2009.

Comitti, V. S.; Shikida, C. D.; Figueiredo, E. Days of Zipf and Covid? Looking for evidence of Zipf's law in the infected Brazil. 2022. Available at SSRN: https://ssrn.com/abstract=4060272.

Crafts, N.; Woltjer, P. Growth accounting in economic history: findings, lessons and new directions. *Journal of Economic Surveys*, Wiley Online Library, v. 35, n. 3, p. 670–696, 2021.

Edmond, C.; Midrigan, V.; Xu, D. Y. Competition, markups, and the gains from international trade. *American Economic Review*, v. 105, n. 10, p. 3183–3221, 2015.

Edmond, C.; Midrigan, V.; Xu, D. Y. How costly are markups? *Journal of Political Economy*, The University of Chicago Press, 2022.

Feenstra, R. C.; Inklaar, R.; Timmer, M. P. The next generation of the Penn World Table. *American Economic Review*, v. 105, n. 10, p. 3150–3182, 2015.

Fernald, J. G. Productivity and potential output before, during, and after the great recession. *NBER Macroeconomics Annual*, University of Chicago Press Chicago, IL, v. 29, n. 1, p. 1–51, 2015.

Fujiwara, Y. Zipf law in firms bankruptcy. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 337, n. 1-2, p. 219–230, 2004.

Fujiwara, Y.; Guilmi, C. D.; Aoyama, H.; Gallegati, M.; Souma, W. Do Pareto–Zipf and Gibrat laws hold true? an analysis with European firms. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 335, n. 1-2, p. 197–216, 2004.

Gabaix, X. Zipf's law for cities: An explanation. *The Quarterly Journal of Economics*, Oxford University Press, v. 114, n. 3, p. 739–767, 1999.

Gabaix, X. Power laws in economics and finance. *Annual Review of Economics*, Annual Reviews, v. 1, n. 1, p. 255–293, 2009.

Gabaix, X. Power laws in economics: An introduction. *Journal of Economic Perspectives*, American Economic Association, v. 30, n. 1, p. 185–206, 2016.

Gabaix, X.; Ibragimov, R. Rank-1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 29, n. 1, p. 24–39, 2011.

Gabaix, X.; Ioannides, Y. M. The evolution of city size distributions. In: North-Holland. *Handbook of Regional and Urban Economics.* : North-Holland, 2004. v. 4, Chap. 53, p. 2341–2378.

Gabaix, X.; Landier, A. Why has CEO pay increased so much? *The Quarterly Journal of Economics*, Oxford University Press, v. 123, n. 1, p. 49–100, 2008.

Giovanni, J. D.; Levchenko, A. A. Firm entry, trade, and welfare in zipf's world. *Journal of International Economics*, Elsevier, v. 89, n. 2, p. 283–296, 2013.

Giovanni, J. D.; Levchenko, A. A.; Ranciere, R. Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics*, Elsevier, v. 85, n. 1, p. 42–52, 2011.

Goerlich, F. J. A simple and efficient test for the pareto law. *Empirical Economics*, Springer, v. 45, n. 3, p. 1367–1381, 2013.

Gollin, D. Getting income shares right. *Journal of Political Economy*, The University of Chicago Press, v. 110, n. 2, p. 458–474, 2002.

Gordon, R. J. Does the "new economy" measure up to the great inventions of the past? *Journal of Economic Perspectives*, American Economic Association, v. 14, n. 4, p. 49–74, 2000.

Hall, R. E. The relation between price and marginal cost in us industry. *Journal of Political Economy*, The University of Chicago Press, v. 96, n. 5, p. 921–947, 1988.

Hall, R. E.; Jorgenson, D. W. Tax policy and investment behavior. *The American Economic Review*, JSTOR, v. 57, n. 3, p. 391–414, 1967.

Hansen, B. E. *Econometrics*. 2020.

Hill, B. M. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, JSTOR, v. 3, n. 5, p. 1163–1174, 1975.

Hsieh, C.-T.; Klenow, P. J. Misallocation and manufacturing TFP in China and India. *The Quarterly journal of economics*, Oxford University Press, v. 124, n. 4, p. 1403–1448, 2009.

Hulten, C. R. Growth accounting. In: Hall, B. H.; Rosenberg, N. (Ed.). *Handbook of the Economics of Innovation*. : Elsevier, 2010. v. 2, p. 987–1031.

Jones, C. I. The facts of economic growth. In: Taylor, J. B.; Uhlig, H. (Ed.). *Handbook of Macroeconomics*. : Elsevier, 2016. v. 2, Chap. 1, p. 3–69.

Jr, N. J. M.; Ribeiro, M. B. Zipf law for Brazilian cities. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 367, p. 441–448, 2006.

Júnior, J. R. d. C. S.; Cornelio, F. M. Estoque de capital fixo no brasil: Séries desagregadas anuais, trimestrais e mensais. *Ipea Working Paper*, n. 2580, 2020.

Justo, W. R. Zipf's law and the Gibrat's law: what do the facts have to say about the brazilian cities? *Journal of Finance and Economics*, v. 2, n. 5, p. 136–144, 2014.

Klenow, P. J.; Rodríguez-Clare, A. The neoclassical revival in growth economics: Has it gone too far? *NBER macroeconomics annual*, MIT Press, v. 12, p. 73–114, 1997.

Kondo, I. O.; Lewis, L. T.; Stella, A. Heavy tailed but not Zipf: Firm and establishment size in the United States. *Journal of Applied Econometrics*, Wiley Online Library, 2023.

Levinsohn, J.; Petrin, A. Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, Wiley-Blackwell, v. 70, n. 2, p. 317–341, 2003.

Loecker, J. D.; Eeckhout, J. Global market power. *NBER Working Paper Series*, n. 24768, 2018.

Loecker, J. D.; Eeckhout, J.; Mongey, S. Quantifying market power and business dynamism in the macroeconomy. *NBER Working Paper Series*, n. 28761, 2021.

Loecker, J. D.; Eeckhout, J.; Unger, G. The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, Oxford University Press, v. 135, n. 2, p. 561–644, 2020.

Loecker, J. D.; Warzynski, F. Markups and firm-level export status. *American Economic Review*, v. 102, n. 6, p. 2437–2471, 2012.

Luttmer, E. G. J. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, Oxford University Press, v. 122, n. 3, p. 1103–1144, 2007.

Melitz, M. J. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, Wiley Online Library, v. 71, n. 6, p. 1695–1725, 2003.

Melitz, M. J.; Redding, S. J. Heterogeneous firms and trade. *Handbook of International Economics*, Elsevier, v. 4, p. 1–54, 2015.

Neary, J. P. Two and a half theories of trade. *The World Economy*, Wiley Online Library, v. 33, n. 1, p. 1–19, 2010.

Okuyama, K.; Takayasu, M.; Takayasu, H. Zipf's law in income distribution of companies. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 269, n. 1, p. 125–131, 1999.

Olley, S.; Pakes, A. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, Wiley Online Library, v. 64, n. 6, p. 1263–1297, 1996.

Peters, M. Heterogeneous markups, growth, and endogenous misallocation. *Econometrica*, Wiley Online Library, v. 88, n. 5, p. 2037–2073, 2020.

Petrin, A.; Levinsohn, J. Measuring aggregate productivity growth using plant-level data. *The Rand Journal of Economics*, Wiley Online Library, v. 43, n. 4, p. 705–725, 2012.

Prescott, E. C. Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity. *International Economic Review*, JSTOR, v. 39, n. 3, p. 525–551, 1998.

Resende, M. Lei de Gibrat na indústria brasileira: Evidência empırica. *Revista EconomiA*, 2004.

Resende, M.; Cardoso, V. Firm size distribution and power laws in Brazil: Some empirical evidence. *IE-UFRJ Discussion Papers*, n. 004, 2022.

Restuccia, D.; Rogerson, R. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, Elsevier, v. 11, n. 4, p. 707–720, 2008.

Restuccia, D.; Rogerson, R. Misallocation and productivity. *Review of Economic Dynamics*, Elsevier, v. 16, n. 1, p. 1–10, 2013.

Santos, T. T. O. A general characterization of the capital cost and the natural interest rate: an application for Brazil. *Central Bank of Brazil Working Paper Series*, n. 524, 2020.

Schluter, C. Top incomes, heavy tails, and rank-size regressions. *Econometrics*, MDPI, v. 6, n. 1, p. 1–16, 2018.

Schluter, C. On Zipf's law and the bias of Zipf regressions. *Empirical Economics*, Springer, v. 61, n. 2, p. 529–548, 2021.

Silva, S. D.; Matsushita, R.; Giglio, R.; Massena, G. Granularity of the top 1,000 Brazilian companies. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 512, p. 68–73, 2018.

Stanley, M. H. et al. Zipf plots and the size distribution of firms. *Economics Letters*, Elsevier, v. 49, n. 4, p. 453–457, 1995.

Syverson, C. Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, v. 33, n. 3, p. 23–43, 2019.

Traina, J. Is aggregate market power increasing? production trends using financial statement. *Stigler Center New Working Paper Series*, n. 17, 2018.

Urzúa, C. M. A simple and efficient test for Zipf's law. *Economics Letters*, Elsevier, v. 66, n. 3, p. 257–260, 2000.

Urzúa, C. M. Testing for Zipf's law: A common pitfall. *Economics Letters*, Elsevier, v. 112, n. 3, p. 254–255, 2011.

Urzúa, C. M. A simple test for power-law behavior. *The Stata Journal*, SAGE Publications, v. 20, n. 3, p. 604–612, 2020.

Vasconcelos, R. Misallocation in the Brazilian manufacturing sector. *Brazilian Review of Econometrics*, v. 37, n. 2, p. 191–232, 2017.

Virkar, Y.; Clauset, A. Power-law distributions in binned empirical data. *The Annals of Applied Statistics*, Institute of Mathematical Statistics, v. 8, n. 1, p. 89–119, 2014.

Vries, G. J. D. Productivity in a distorted market: The case of Brazil's retail sector. *Review of Income and Wealth*, Wiley Online Library, v. 60, n. 3, p. 499–524, 2014.

Vuong, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, Wiley Online Library, v. 57, n. 2, p. 307–333, 1989.

Wang, O.; Werning, I. Dynamic oligopoly and price stickiness. *American Economic Review*, v. 112, n. 8, p. 2815–49, 2022.

Zipf, G. K. Human behavior and the principle of least effort: An introduction to human eoclogy. Addison-Wesley Press, 1949.